FACULTY OF PSYCHOLOGY AND
EDUCATIONAL SCIENCES
BRAIN AND COGNITION

Deriving affective meaning from connectivity in the mental lexicon

Bram Van Rensbergen

Doctoral Thesis offered to obtain
the degree of Doctor of Psychology

Supervisor: Prof. Dr. Gert Storms
Co-supervisor: Dr. Simon De Deyne

2016

Bram Van Rensbergen. **Deriving affective meaning from connectivity in the mental lexicon.**

Dissertation submitted in order to obtain the degree of Doctor in Psychology, 2015.

Promotor Prof. Dr. Gert Storms Co-promotor Dr. Simon De Deyne

The mental lexicon, the structure reflecting a person's knowledge of words, contains not only information about a word's denotation, pronunciation, and part of speech, but also about the word's connotation or affective meaning. While there has been considerable research on the mental lexicon, most has focused on denotational and linguistic aspects of words; connotation has not received as much attention. In this dissertation, we will examine the relation between affective meaning and connectivity in the mental lexicon.

In a first empirical study (Chapter 2), we use a large word association dataset to investigate whether connected words share affective attributes. We find that words tend to be connected to words of similar valence, arousal, dominance, and concreteness. Considering this, it seems reasonable to assume that we can obtain information about a word's affective meaning from the words it is connected to. We examine this possibility in three further chapters.

In Chapter 3, we outline a method to predict the valence, arousal, and dominance of words, using their connectivity towards words for which these values are already known. We find that obtained predictions show very high correlations to human ratings.

In Chapter 4, we follow a similar approach to estimate the correspondence of words towards the Big Five personality dimensions: openness, conscientiousness, extraversion, agreeableness, and neuroticism. We use these estimates to code the responses of a free-format personality test, in which participants describe their own personality using any ten words. We find that the resulting personality profiles show a strong correspondence to profiles obtained by having trained psychologists code responses.

Finally, in Chapter 5, we investigate the possibility of measuring brand personality, the human characteristics associated with a brand, by examining the connectivity of the associations people make towards a brand. We test this for a number of well-known brands, and find that the resulting brand personality indices show a mixed correspondence to human ratings, with correlations that are high for some dimensions but low and nonsignificant for others.

Bram Van Rensbergen. **Affectieve betekenis afleiden uit connectiviteit in het mentale lexicon.**

Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Psychologie, 2015.

Promotor Prof. Dr. Gert Storms Co-promotor Dr. Simon De Deyne

---

Het mentale lexicon, de structuur die iemands kennis over woorden representeert, bevat niet alleen informatie over de denotatie, uitspraak, en lexicale categorie van woorden, maar ook over connotatie of affectieve betekenis. Het meeste onderzoek over het mentale lexicon is toegespitst op de denotationele en linguïstische aspecten van woorden; in vergelijking heeft connotatie aanzienlijk minder aandacht gekregen. In deze dissertatie bestuderen we de relatie tussen affectieve betekenis en connectiviteit in het mentale lexicon.

In een eerste empirische studie (Hoofdstuk 2) gebruiken we woordassociaties om te kijken of woorden die mensen als gerelateerd zien affectieve eigenschappen gemeen hebben. We vinden dat woorden een neiging tonen om verbonden te zijn met woorden met een gelijkaardige valentie, activiteit, dominantie, en concreetheid. Het lijkt dus plausibel dat we informatie over de affectieve betekenis van een woord kunnen afleiden van de woorden waarmee het verbonden is; we bekijken in hoeverre dit mogelijk is in drie verdere hoofdstukken.

In Hoofdstuk 3 schetsen we een methode om de valentie, activiteit, en dominantie van woorden te voorspellen, gebruik makende van hun connectiviteit met woorden waarvoor deze waardes al gekend zijn. We vinden dat de voorspelde waardes zeer hoge correlaties vertonen met menselijke ratings.

In Hoofdstuk 4 volgen we een gelijkaardige aanpak om te voorspellen in hoeverre woorden overeenstemmen met de Big Five persoonlijkheidsdimensies: intellectuele autonomie, ordelijkheid, extraversie, mildheid, en emotionele stabiliteit. We gebruiken deze voorspellingen om de antwoorden op een vrije zelfbeschrijving persoonlijkheidstest te coderen; in deze test beschrijven deelnemers hun persoonlijkheid met tien woorden naar keuze. We vinden dat de resulterende persoonlijkheidsprofielen sterk overeenkomen met profielen verkregen door deze test te laten coderen door getrainde psychologen.

In Hoofstuk 5, ten slotte, bekijken we of we merkpersoonlijkheid, de menselijke karakteristieken die geassocieerd worden met een merk, kunnen meten door te kijken naar de connectiviteit van woorden waarmee een merk geassocieerd wordt. We testen dit voor een aantal bekende merken, en vinden dat de verkregen merkpersoonlijkheidsprofielen een wisselvallige correspondentie toont met menselijke ratings: bij enkele dimensies is de correlatie zeer hoog, maar bij anderen is die laag en niet significant.

Bedankt

Aan Gert,
voor de deur die altijd open stond
voor de vrijheid en het vertrouwen
voor de raad en je relativeringstalent

Aan Simon,
voor de motivatie en de inspiratie
voor je eindeloze enthousiasme over woordassociaties
voor je uitgebreide feedback, telkens opnieuw

Aan Tom,
voor altijd open te staan voor vragen
voor het advies over statistiek en R

Aan Eef, Wouter, Steven, Loes, Farah, Anne, en Hendrik
voor het delen in vreugde wanneer alles goed liep
voor het delen in miserie wanneer dat niet zo was
voor de gezellige lunches, elke dag opnieuw

Aan familie en vrienden, je weet wie jullie zijn
voor het helpen stoom aflaten
voor de emotionele steun
voor het begrijpen dat 'psychologie' een acceptabel antwoord is op de vraag 'wat onderzoek je'

En zeker ook aan Katrin
zonder jou was het nooit gelukt

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Language is one of the most important aspects of human life. In order to understand language, one needs to understand the words it is made out of. This information is generally considered to be stored in the mental lexicon, the structure containing a person's knowledge of words. This lexicon contains not only the denotation of words (a *dog* is an animal), but also affective meaning (*flowers* are nice), pronunciation (*balloon* is pronounced bə-lo͞onˈ), part of speech (*house* is a noun), and so forth (Jackendoff, 2002).

There has been considerable research on the mental lexicon (for an overview, see Aitchison, 2003). Most of this has focused on denotational and linguistic aspects; affective meaning has not received as much attention, even though it is becoming increasingly clear that affective information plays an essential role both in theories of semantics (De Houwer & Randell, 2004; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Vigliocco, Meteyard, Andrews, & Kousta, 2009) and more generally, in accounts of cognitive systems (Dolan, 2002; Pessoa, 2008; Phelps, 2006).

In this dissertation, we will investigate the relation between affective meaning and connectivity in the mental lexicon. We will start by examining whether connected words share affective meaning; for example, whether words that are considered positive tend to be connected to other positive words. We then examine whether we can use this relation to predict the affective meaning of words, from the words with which they are connected.

To do so, we will need to obtain a measure of the mental lexicon. At this point, we should clarify that there is no such thing as *the* mental lexicon; every single person has his or her own version, reflecting their personal experience and knowledge (Aitchison, 2003). Importantly, though, people show a remarkable consistency in terms of which words they deem to be related (Aitchison, 2003; De Deyne, Navarro, Perfors, & Storms, 2012; Postman & Keppel, 1970), which is perhaps not that surprising as one of the goals of language is communication. As such, it is possible to combine the data of many persons to create an aggregate measure of the mental lexicon, representing a shared cultural artifact.

As we are interested in affective meaning, we want to make sure our measure of the lexicon fully captures these aspects. Subjective meaning is not easy to gauge, as it is something personal and often subconscious (Szalay & Deese, 1978). To make matters worse, measures of opinion or attitude tend to be strongly influenced by social desirability effects and other forms of response bias (see for example Edwards, 1957; Furnham, 1986). One promising approach to investigate subjective meaning in the mental lexicon is using word association data (Deese, 1965).

1.1 Word associations

In a word association task, a participant writes down the first words that come to mind spontaneously after reading a cue word. The probability that a cue elicits a certain response can then be considered a measure of the associative strength between that cue and response in the mental lexicon (Cramer, 1968; De Deyne, Navarro, & Storms, 2013; Mollin, 2009; Nelson, McEvoy, & Dennis, 2000; Nelson, McEvoy, & Schreiber, 2004; for further evidence see also Gallagher & Palermo, 1967; Griffiths, Steyvers, & Tenenbaum, 2007; Nelson, McKinney, Gee, & Janczura, 1998; Wicklund, Palermo, & Jenkins, 1965). This direct connection to what goes on in the mind makes the word association task an excellent measure of affective meaning, as it often grants access to subconscious attitudes that cannot always be verbalized in response to direct questions (Szalay & Deese, 1978), and allows little chance for participants to consciously monitor their responses (Deese, 1965; Szalay & Deese, 1978).

When using word association data to investigate the relation between two words, it is important to consider not only their direct connection (i.e., how often one word is given as an association in response to the other), but also whether they occur in similar contexts; for example, whether they elicit the same associates or are given as associations to the same words (e.g., De Deyne & Storms, 2015, p. 466; Deese, 1965; Maki, 2007, Szalay & Deese, 1978). Using this approach, word association data can capture relations between words that are judged as related in direct comparisons, yet are not frequently given as an association to one another (De Deyne et al., 2013; De Deyne, Verheyen, & Storms, 2015; Gravino, Servedio, Barrat, & Loreto, 2012).

Word associations are not the only method of constructing a representation of the mental lexicon. Other options include computational models based on text corpora such as latent semantic analysis (Landauer & Dumais, 1997) or topic models (Steyvers, & Tenenbaum, 2007), or networks constructed using linguistic expert knowledge such as WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990). Compared with these alternatives, the associative technique offers a number of advantages.

First, as we have already seen, word associations are very resistant against response bias and other forms of cognitive monitoring, making them an excellent measure of the subjective aspects of the mental lexicon.

Moreover, word associations combine the best of two worlds. Similar to models based on text corpora, they capture lexical co-occurrence relations (Hahn & Sivley, 2011; Schrijnemakers &

Raaijmakers, 1997; Wettler, Rapp, & Sedlmeier, 2005); and like approaches based on expert knowledge, they capture meaningful relations that cannot be explained by lexical co-occurrence alone (McRae, Khalkhali, & Hare, 2011; Mollin, 2009). For example, word association data reflects multiple meanings of ambiguous words, and can help indicate which meanings are considered dominant. Additionally, mentally central properties such as color or shape (e.g., *banana–yellow* or *ball–round*) are well represented in association data, while they are somewhat uncommon in text corpora. This is likely because word associations and text corpora represent information of a different nature: the goal of written text is to communicate some message efficiently, and information that is known to both parties is often omitted (Grice, 1975; Sperber & Wilson, 2005); word associations, in contrast, are simply a non-propositional expression of thought, and are usually free from pragmatics and intent (Deese, 1965; McRae, Khalkhali, & Hare, 2011; Mollin, 2009; Szalay & Deese, 1978).

A more general advantage of word associations is that they can be used to easily investigate any type of cue, irrespective of its lexical or semantic properties, and even regardless of form of presentation (as written text, auditory, in the form of images, ...). Most existing research on the mental lexicon investigates semantic categories (e.g., mammals or weapons), adjectives (e.g., Gross, Fischer, & Miller, 1989), or verbs (e.g., Gentner, 1978), yet the lexicon includes all conceivable types of words, including proper names (e.g., Paris, Coca Cola, or Leonardo da Vinci), adverbs, prepositions, and so forth. The associative method is a flexible approach to investigating these words.

As we have hinted at above, when studying meaning with word associations it is important to not limit yourself to single connections between words, but rather to take into account a word's entire response distribution (Deese, 1968; Szalay & Deese, 1978). With this approach it becomes feasible to examine the structure of the mental lexicon as a whole, by representing it as a semantic network where nodes correspond to words and connections indicate a meaningful relation between them (Collins & Loftus, 1975; Collins & Quillian, 1969; De Deyne & Storms, 2015).

One caveat of using word associations to investigate the structure of the lexicon is that it requires access to a very large dataset; smaller word association datasets usually lack the heterogeneity in responses to fully represent the mental lexicon, especially when they were created by asking for a single association per cue (Aitchison, 2003). Creating a sufficiently large word association database is a considerable undertaking, as it requires data from an extensive number of participants. Fortunately, datasets of this nature already in exist in many languages. In this dissertation, we will use the Dutch Small World of Words project[1], a recent corpus that contains over five million associations collected in response to roughly 16,000 cue words (De Deyne et al., 2013). Each cue was present to about 100 participants, who responded with up to three associations per cue. Compared with other word association datasets, the Small World of Words database is considerably larger, and as such covers a much larger part of the human lexicon. Indeed, the 16,000 cue words represent over 90% of the words encountered in written text (De Deyne et al., 2013). An additional advantage of this dataset is that by asking for three associations per cue rather than just one, it encodes much more heterogeneous responses, including non-dominant and weaker associations (De Deyne & Storms, 2008).

---

[1] See `www.smallworldofwords.com`.

1.2 A quick run-through

While it is generally accepted that the mental lexicon encodes subjective information, there has been little research into the contribution of an affective component to the structure of the lexicon. In this dissertation, we will address this issue by examining the relation between affective meaning and connectivity in the mental lexicon.

In Chapter 2, we examine whether connected words share affective values. We find that words tend to be connected to words of similar valence, arousal, dominance, and concreteness. Furthermore, we find that this finding extends to concreteness, a non-affective factor that has been linked previously to affective meaning (e.g., Vigliocco, Meteyard, Andrews, & Kousta, 2009).

In the subsequent three chapters, we investigate whether we can use this relation to predict affective word values based on words' connectivity in the mental lexicon.

In Chapter 3, we predict the valence, arousal, and dominance of a large number of words, using their connectivity towards words for which these values are already known. We find that obtained predictions show very high correlations to human ratings.

Chapter 4 follows a similar approach to estimate the correspondence of words towards the Big Five personality dimensions: openness, conscientiousness, extraversion, agreeableness, and neuroticism. We use these estimates to code the responses of a free-format personality test, in which participants describe their own personality using any ten words. The resulting personality profiles correspond strongly to profiles obtained by having trained psychologists code responses.

Finally, in Chapter 5, we use word association data to measure brand personality, the human characteristics associated with a brand. We collect associations towards a number of well-known brands, and use the connectivity of obtained responses to predict the responsibility, activity, aggressiveness, simplicity, and emotionality of each brand. We find that the resulting brand personality indices show a moderate correspondence to direct ratings.

Note that each of these chapters corresponds to a self-contained manuscript; as a result, a certain degree of overlap is inevitable. Moreover, as the manuscripts were written for publication in academic journals in different fields of research, each will approach the topic at hand from a different point of view.

References

Aitchison, J. (2003). *Words in the Mind: An Introduction to the Mental Lexicon*. Hoboken, NJ: John Wiley & Sons.

Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407.

Collins, A., & Quillian, M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247.

Cramer, P. (1968). *Word association*. New York, NY: Academic Press.

De Deyne, S., Navarro, D. J., Perfors, A., & Storms, G. (2012). Strong structure in weak semantic similarity: A graph based account. In *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1464–1469). Austin, TX: Cognitive Science Society.

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–98. doi:10.3758/s13428-012-0260-7

De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, *40*(1), 198–205. doi:10.3758/BRM.40.1.198

De Deyne, S., & Storms, G. (2015). Word Associations. In J. R. Taylor (Ed.), *The Oxford Handbook of the Word* (pp. 465–480). Oxford, U.K.: Oxford University Press.

De Deyne, S., Verheyen, S., & Storms, G. (2015). The role of corpus size and syntax in deriving lexico-semantic representations for a wide range of concepts. *Quarterly Journal of Experimental Psychology*, *68*(8), 1643–1664. doi:10.1080/17470218.2014.994098

De Houwer, J., & Randell, T. (2004). Robust affective priming effects in a conditional pronunciation task: Evidence for the semantic representation of evaluative information. *Cognition & Emotion*, *18*(2), 251–264. doi:10.1080/02699930341000022

Deese, J. (1965). *The structure of associations in language and thought*. Baltimore, MD: Johns Hopkins University Press.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Ft Worth, TX: Dryden Press.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, *7*(3), 385–400. doi:10.1016/0191-8869(86)90014-0

Gallagher, J., & Palermo, D. (1967). The effect of type associative relation on word recognition times. *Child Development*, *38*(3), 849–855. doi:10.2307/1127262

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development*, *49*(4), 988–998.

Gravino, P., Servedio, V. D. P., Barrat, A., & Loreto, V. (2012). Complex structures and semantics in free word association. *Advances in Complex Systems*, *15*(03n04). doi:10.1142/S0219525912500543

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244. doi:10.1037/0033-295X.114.2.211

Hahn, L. W., & Sivley, R. M. (2011). Entropy, semantic relatedness and proximity. *Behavior Research Methods*, *43*(3), 746–60. doi:10.3758/s13428-011-0087-7

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. New York, NY: Oxford University Press.

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34. doi:10.1037/a0021446

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. doi:10.1037/0033-295X.104.2.211

McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and associative relations in adolescents and young adults: Examining a tenuous dichotomy. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), *The adolescent brain: Learning, reasoning, and decision making* (pp. 39–66). Washington, DC: American Psychological Association. doi:10.1037/13493-002

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*(4), 235–244.

Mollin, S. (2009). Combining corpus linguistic and psychological data on word co-occurrences: Corpus collocates versus word associations. *Corpus Linguistics and Linguistic Theory*, *5*(2), 175–200. doi:10.1515/CLLT.2009.008

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition*, *28*(6), 887–99.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–7. doi:10.3758/BF03195588

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, *105*(2), 299–324. doi:10.1037/0033-295X.105.2.299

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, *9*(2), 148–158. doi:10.1038/nrn2317

Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology*, *57*, 27–53. doi:10.1146/annurev.psych.56.091103.070234

Schrijnemakers, J., & Raaijmakers, J. (1997). Adding new word associations to semantic memory: Evidence for two interactive learning components. *Acta Psychologica*, *96*(1), 103–132.

Sperber, D., & Wilson, D. (2005). Pragmatics. In F. Jackson & M. Smith (Eds.), *Oxford Handbook of Contemporary Analytic Philosophy* (pp. 468–501). Oxford, U.K.: Oxford University Press.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. doi:10.1515/LANGCOG.2009.011

Wettler, M., Rapp, R., & Sedlmeier, P. (2005). Free Word Associations Correspond to Contiguities Between Words in Texts. *Journal of Quantitative Linguistics*, *12*(2-3), 111–122. doi:10.1080/09296170500172403

Wicklund, D., Palermo, D., & Jenkins, J. (1965). Associative clustering in the recall of children as a function of verbal association strength. *Journal of Experimental Child Psychology*, *2*(1), 58–66. doi:10.1016/0022-0965(65)90015-9

Chapter 2

Assortativity in the mental lexicon

This chapter was published as:

Van Rensbergen, B., Storms, G., & De Deyne, S. (2015). Examining assortativity in the mental lexicon: Evidence from word associations. *Psychonomic Bulletin & Review, 22*(6), 1717-1724. doi:10.3758/s13423-015-0832-5

Abstract

Words are characterized by a variety of lexical and psychological properties, such as their part of speech, word-frequency, concreteness, or affectivity. In this study, we examine how these properties relate to a word's connectivity in the mental lexicon, the structure containing a person's knowledge of words. In particular, we examine the extent to which these properties display assortative mixing, that is, the extent to which words in the lexicon are more likely to be connected to words that share these properties. We investigate three types of word properties: (1) subjective word covariates: valence, dominance, arousal, and concreteness, (2) lexical information: part of speech, and (3) distributional word properties: age-of-acquisition, word frequency, and contextual diversity. We assess which of these factors exhibit assortativity using a word association task, where the probability of producing a certain response to a cue is a measure of the associative strength between the cue and response in the mental lexicon. Our results show that the extent to which these aspects exhibit assortativity varies considerably, with a high cue-response correspondence on valence, dominance, arousal, concreteness, and part of speech, indicating that these factors correspond to the words people deem as related. In contrast, we find that cues and responses show only little correspondence on word frequency, contextual diversity, and age-of-acquisition, indicating that, compared with subjective and lexical word covariates, distributional properties exhibit only little assortativity in the mental lexicon. Possible theoretical accounts and implications of these findings are discussed.

2.1 Introduction

The mental lexicon, defined by Jackendoff (2002) as the store of words in long-term memory from which the grammar constructs phrases and sentences, contains information such as part of speech (*house* is a noun), denotation (a *dog* is an animal), pronunciation (*balloon* is pronounced bə-lo͞onˈ), affective meaning (*cake* is something I like), and so forth. When studying aspects of word meaning, the mental lexicon is sometimes portrayed as a semantic network, in which nodes correspond to words and connections indicate a meaningful relation between them (Collins & Loftus, 1975; Collins & Quillian, 1969).

While connections between concepts often reflect semantic relationships (e.g., synonymy, hyponymy, meronomy, …; see Murphy, 2003), research suggests that the properties of a word itself correlate with connectivity as well. In particular, a small corpus of studies indicates that the probability that two words are connected correlates with the presence of similar lexical or psychological properties. In network terms, this tendency for connected nodes to exhibit similar covariates is called *assortativity* or *assortative mixing*[2] (e.g., Newman, 2010; Vitevitch, 2008; Vitevitch, Chan, & Goldstein, 2014).

To study assortative mixing, word association data are often used. In a word association task, the probability of producing a certain response to a cue is a measure of the associative strength between the cue and response in the lexicon (De Deyne, Navarro, & Storms, 2015; Nelson,

---

[2] Note that this term indicates a specific type of mixing, as it only refers to the tendency of nodes to attach to others that are similar in some way. The opposite situation, where attachment is driven by dissimilarity, is referred to as disassortative mixing. Since throughout this study no evidence is found for disassortative mixing, we only discuss the positive case.

McEvoy, & Schreiber, 2004). As such, a cue-response correspondence on some factor would be indicative of that factor displaying assortative mixing in the mental lexicon. Using this approach, word association research has identified several factors that exhibit assortativity, that is, several properties that tend to be shared between connected concepts.

Firstly, there is evidence for assortative mixing by syntax: in a word association task, cues tend to elicit responses with the same syntactic properties (e.g., Cramer, 1968; Deese, 1962, 1965). These results are corroborated by the finding that processing an utterance with a specific syntactic form facilitates processing utterances with a similar syntax (a phenomenon named syntactic priming, see e.g., Bock, 1986; or Pickering & Branigan, 1998, 1999), by the finding that word selection errors frequently preserve part of speech (Hotopf, 1980), and by noun- or verb-specific deficits in patient studies (Mätzig, Druks, Masterson, & Vigliocco, 2009).

There is also evidence that valence (i.e., how positive a word is considered, cfr. Osgood, Suci, & Tannenbaum, 1957) exhibits assortativity, as research shows a positive cue-response correlation on this dimension (Cramer, 1968; Pollio, 1964; Staats & Staats, 1959), and activation of a specific evaluative attitude (e.g., *good*) facilitates processing of information that shares that evaluation (a concept called affective priming: see Klauer, 1997, for an overview).

Similarly, word association studies show evidence for assortative mixing by dominance (whether a word refers to a strong or dominant concept, e.g., *power*) and arousal (whether a word refers to an active or aroused concept, e.g., *explosion*), again evidenced by positive cue-response correlations on these aspects (Pollio, 1964; Staats & Staats, 1959).

Finally, research on concreteness (the extent to which words are imageable i.e., refer to something perceptible) suggests this factor may exhibit assortativity as well, as processing a concept with a specific degree of concreteness facilitates processing of concepts with similar imageability (Bleasdale, 1987).

Research on the structure of the mental lexicon has not been limited to assessments of assortativity. A separate line of inquiry has focused on uncovering which word properties contribute to the overall number of connections a word has, that is, what aspects determine which nodes are highly connected or central in the mental lexicon, and which are not. Some of this research examined the same word properties described above, observing, for example, that words with a high valence show increased connectivity (Cramer, 1968; Johnson & Lim, 1964; Matlin & Stang, 1978; Pollio, 1964), as do highly imageable words (de Groot, 1989). Other researchers investigated the role of statistical word properties that are not related directly to meaning, but are inferred from the environment in which a word is acquired. They find that concepts that are learned at a young age show higher network connectivity (Barabási & Albert, 1999; Steyvers & Tenenbaum, 2005), and that a person's exposure to a particular word is involved as well: words with a high word frequency show higher network connectivity (Steyvers & Tenenbaum, 2005), as do words with a high contextual diversity (the number of different contexts in which a word is seen; Hills, Maouene, Riordan, & Smith, 2010). Clearly, these distributional word properties are linked to the structure of the mental lexicon, yet to our knowledge, no research has assessed whether they exhibit assortative mixing, which considers similarity of connected concepts, and is distinct from a relation between these factors and overall connectivity.

2.1.1 The current study

As indicated above, a number of studies have identified several word covariates that display assortativity in the mental lexicon: part of speech, valence, dominance, arousal, and concreteness. Yet, none of these studies have investigated these factors simultaneously, which makes it very hard to evaluate whether they exert an independent contribution. Potentially, these factors depend on one another; it is conceivable, for example, that after controlling for one factor, the effects of some other factor(s) disappear. In the same vein, the lack of common ground between these studies makes it hard to estimate the relative importance of each factor.

A second problem is that part of the research that looked into these factors made use of very small sample sizes, mostly because of technical limitations of their time, making generalizations towards the entire mental lexicon somewhat unfeasible. For example, the study of Staats & Staats (1959) was based on 10 words, and the study of Pollio (1964) comprised 52 words; these small stimulus sets are likely to misrepresent the variability captured by a combination of the investigated factors.

In this study, we use word association data to investigate the linguistic and subjective factors that are involved with the configuration of the mental lexicon, by examining the extent to which cue word and their associative responses exhibit similar properties. We investigate part of speech, valence, dominance, arousal, and concreteness, five factors that have previously been established to display assortativity. We will also examine word frequency, contextual diversity, and age-of-acquisition, three aspects that have been found to be involved with the structure of the mental lexicon, but for which assortativity has not yet been assessed.

Our main goal is, then, to (a) establish which of these factors display assortativity in the mental lexicon, (b) investigate their relative contribution, and (c) examine whether these findings uphold for a large variability of cue stimuli.

2.2 Method

2.2.1 Materials

2.2.1.1 Word association corpus.

To derive the associative strength for a large set of items, we made use of the Dutch Small World of Words project, which comprises 3.8 million cue-response pairs (see De Deyne, Navarro, & Storms, 2013, for full details). Briefly, these associations were gathered in response to over 12,571 cues; each cue was presented to 100 participants, who gave up to three responses to a number of cues in a continued word association task.

2.2.1.2 Lexical and psycho-affective variables.

Three norming databases were used to gather lexical and psycho-affective measures of a large set of words. Word frequency, contextual diversity, and syntactic form (part of speech) for 437,000 Dutch words was obtained from Keuleers, Brysbaert, and New (2010). Word frequency was derived from the raw word count in the subtitles of 8,070 films and television show episodes, contextual diversity was based on the number of films or episodes a word occurred in, and part of speech was estimated using an integrated Dutch morphosyntactic analyzer and part of speech tagger (Tadpole: Van Den Bosch, Busser, Canisius, & Daelemans, 2007).

Age-of-acquisition estimates and concreteness ratings for 30,000 Dutch words were taken from the dataset by Brysbaert, Stevens, De

Deyne, Voorspoels, and Storms (2014). Age-of-acquisition was estimated in years, while concreteness was rated on a 5-point Likert scale, where a value of 1 corresponded to 'very abstract', and a value of 5 to 'very concrete'.

Valence, arousal, and dominance ratings for 4,300 Dutch words were available through Moors et al. (2013). Each dimension was rated on a 7-point Likert scale, where a value of 1 corresponded to 'very negative/unpleasant', 'very passive/calm', and 'very weak/submissive', respectively, and a value of 7 to 'very positive/pleasant, 'very active/aroused', and 'very strong/dominant'. Cues in this database were selected from various sources and consisted of mostly nouns, adjectives, and verbs.

2.2.2 Procedure

Out of the 3.8 million cue-response pairs in the Dutch Small World of Words project, 665,461 consist of a cue and response both present in all three norming databases described above. These word pairs contain 4,151 unique words (2472 nouns, 764 verbs, 814 adjectives, and 101 other word types, based on the dominant syntactical role described by Keuleers, Brysbaert, and New, 2010).

2.3 Results

To investigate the extent to which part of speech, valence, arousal, dominance, concreteness, word frequency, contextual diversity, and age-of-acquisition display assortativity in the mental lexicon, we assessed how cues and associative responses correspond on these factors. Our main objective was to inspect correspondence within one dimension, that is, how much of the variance in associative responses' values on some factor is explained by cue values on that factor. A secondary goal was to examine the extent to which the different factors depend on each other.

To this end, we fitted seven multiple linear regression models, each of which predicts response values on one factor using cue values on all seven measures. The relative contribution of each predictor in the regression model was assessed using the metric *lmg* in the R package *relaimpo* (Grömping, 2006), which takes into account predictor collinearity, and handles the issue of predictor order by averaging across all possible orders. The resulting R^2 values are described in Table 2.1.

For affective dimensions, we find that response values are by far best predicted by cue values on that same measure, as one might expect if these aspects display assortativity. Cues and responses correspond most strongly on valence, with cue valence explaining 31% of the variance in response valence. We find a smaller but still considerable cue-response correspondence on arousal, dominance, and concreteness, with cue properties explaining between 15% and 20% of variance in response values.

We find almost no cue-response correspondence on word frequency and contextual diversity, with cue properties explaining at most 1% of variance in response values. Lastly, we find a small effect-size of age-of-acquisition, with cue age-of-acquisition explaining 4% of variance in response age-of-acquisition.

Scatterplots of cue and response values reveal distributions that are somewhat skewed, at least for some of the examined variables (Figure 2.1). As such, it is possible that the cue-response correspondence displayed in Table 2.1 is the result of the distributional properties of the used data, instead of being indicative of assortative mixing. To investigate this alternate explanation, we performed the above regression analysis after permuting the cue-association pairs (so responses are not matched to ‘their’ cue, but to a random cue). This approach yields R^2 values smaller than .001 for all predictors in all seven models, which indicates that the R^2 values reported in Table 2.1 are not a result of the properties of the used dataset, but rather indicate that when presented with a cue, people tend to respond with associations of similar valence, arousal, dominance, and concreteness, and to a small extent, age-of-acquisition. [3]

Finally, to investigate cue-response correspondence on part of speech, we include a part of speech contingency table (Table 2.2). We find that overall, 57.50% of responses match the syntactical role of their corresponding cue. Combining the six smallest categories into one (adverbs, pronouns, prepositions, interjections, determiners, and numerals) allows us to perform a chi-squared test on the contingencies, which indicates that part of speech of responses is significantly related to part of speech of their corresponding cue (X^2 = 82,469, df = 9, p < .001, Cramér’s C = .205).

[3] We also investigated whether this cue-response correspondence is mediated by part of speech. We ran the same multiple linear regression models, this time limited to cue-response pairs where the cue is a noun, a verb, or an adjective. Because the findings were very similar to those derived from the entire dataset, with large effect-sizes for affective and lexical dimensions, and minimal correspondence on distributional properties, they are not repeated here.

Table 2.1

Proportion of variance (> .001) in response values on various psychological and lexical dimensions explained by cue values on those dimensions

	Predictors						
Observed Variable	Cue Valence	Cue Arousal	Cue Dominance	Cue Concr.	Cue Word Freq.	Cue Cont. Div.	Cue AoA
Response Valence	**.31**		.01				
Response Arousal		**.17**	.05	.01			
Response Dominance	.01	.04	**.15**				
Response Concr.				**.20**	.01	.01	.03
Response Word Freq.					**.01**	.01	
Response Cont. Div.					.01	**.01**	
Response AoA	.01			.04			**.04**

Note. n = 665,461. Cells contain R^2 values derived from a multiple linear regression model with all seven predictors, analyzed using the *lmg* metric found in R package *relaimpo* (Grömping, 2006). Concr. = concreteness. Word Freq. = log10 of word frequency per million words. Cont. Div. = log10 of contextual diversity. AoA = age-of-acquisition.

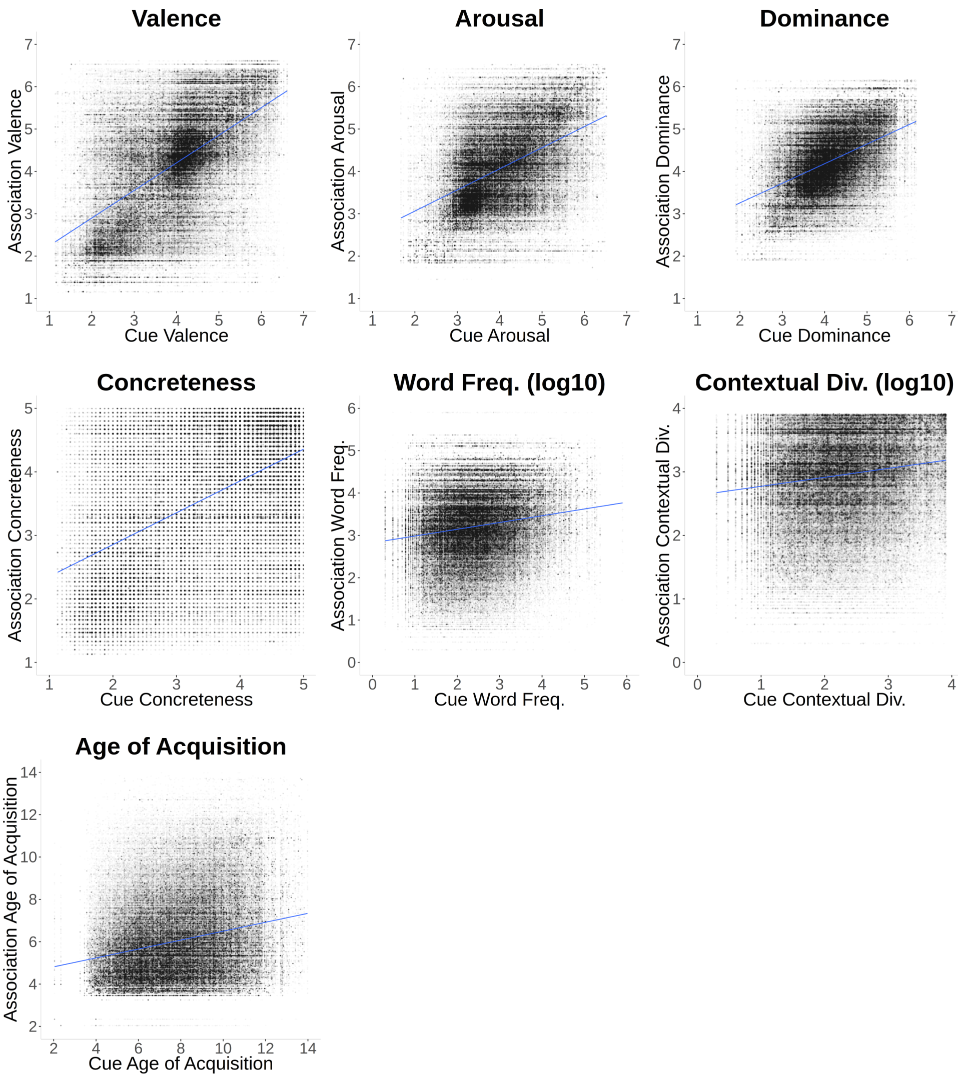

Figure 2.1. Regression lines and scatterplots (with semi-transparent markers) of cue-response correspondence on various psychological and linguistic ratings (*n* = 665,461). Word Freq. = Word Frequency, Contextual Div. = Contextual Diversity.

Table 2.2

Contingency table denoting part of speech of 654,484 cue-response pairs

	Response Part of Speech							
Cue Part of Speech	Noun	Adjective	Verb	Adverb	Pronoun	Preposition	Other	*n*
Noun	**283,415**	81,511	40,926	2,218	2,284	1,415	170	411,939
Adjective	43,680	**65,314**	9,279	1,866	820	206	29	121,194
Verb	65,616	20,293	**27,254**	955	573	214	6	114,911
Adverb	1,437	1,031	320	**242**	118	5	2	3,155
Pronoun	910	507	128	86	**39**	2	0	1,672
Preposition	400	194	60	14	2	**18**	1	689
Other	587	111	107	3	67	15	**34**	924
n	396,045	168,961	78,074	5,384	3,903	1,875	242	654,484

Note: *Other* refers to words tagged as interjection, determiner, or numeral. Cues and responses with unknown or unclear part of speech were omitted (items tagged as *SPEC* in the database of Keuleers, Brysbaert, & New, 2010).

2.4 Discussion

The present research used word association data to assess the assortativity of various linguistic and psycho-affective factors. Using an approach that allows us to compare the relative importance of each factor, we examined valence, arousal, dominance, concreteness, word frequency, contextual diversity, age-of-acquisition, and part of speech.

In investigating cue-response correspondence on these dimensions, we find a very strong assortative effect of valence. This pivotal role of evaluative attitude is in line with existing word association research; for example, Deese (1965) identified valence as the dominant factor in determining which concepts people consider related, and a study of our own found valence to account for over 83% of the variance in a spatial representation of the mental lexicon (De Deyne et al., 2013). The vital importance of evaluative attitude is corroborated in other domains as well, such as in word recognition research (Kuperman, Estes, Brysbaert, & Warriner, 2014), categorization tasks (Niedenthal, Halberstadt, & Innes-Ker, 1999), or affective priming (Klauer, 1997).

We also find a high cue-response correspondence on dominance and arousal, again in line with existing research (Pollio, 1964; Staats & Staats, 1959). This seminal role of the affective dimensions valence, dominance, and arousal is in agreement with the traditional view on semantic meaning. In an attempt to quantify connotative meaning, Osgood and colleagues performed a factor analysis on ratings of concepts on a large number of semantic dimensions (Osgood, Suci, & Tannenbaum, 1957). They found that evaluation (valence), potency (dominance), and activity (arousal) are by far the most powerful aspects in differentiating subjective meaning. Moreover, the importance of these dimensions seems to be near universal, as follow-up studies have replicated these results across dozens of cultures (see Heise, 2010, or Osgood, 1975, for an overview).

In examining concreteness, we find that the level of abstractness of cues is highly predictive of that of its corresponding responses, indicating that this factor, too, is involved with the structure of the mental lexicon. Some research on concreteness-based priming reports similar findings (Bleasdale, 1987), although in general, this factor has received little attention in literature on the mental lexicon. Considering the strong effect we report, inclusion of this factor in future research on the structure of the lexicon might be merited.

Overall, we find that all investigated subjective dimensions show a high cue-response correspondence, indicative of assortative mixing. This is clear evidence for the idea that subjective/affective dimensions are involved with the structure of the mental lexicon, and likely play an important role in shaping chain of thought overall.

We also examined the role of syntactic information. We find that cues tend to elicit associative responses with similar syntactic properties, in concordance with existing research (see Deese, 1965, for an overview). This effect was highly significant; in fact, we find that over half of all associations share the part of speech of their corresponding cue, evidence that syntax exhibits network assortativity as well. We also assessed whether the effects of the psycho-affective and statistical word properties we investigated were mediated by cue part of speech. In comparing results for verb cues, adjective cues, and noun cues, we find some small baseline differences, although all correspondence patterns described above remained true in all three cases.

As described above, existing research also reports evidence for assortative mixing by valence, arousal, dominance, concreteness, and part of speech. Most of these aspects were studied separately; as such, these existing studies cannot rule out the possibility that some of these factors depend on one another. By investigating all aspects simultaneously, we were able to establish that the assortativity effects reported both by us and in this previous literature cannot be explained by any codependence between the different factors; rather, each of these investigated aspects displays assortative mixing independently of any relation to the remaining factors.

A separate concern with existing research on assortativity in the mental lexicon is that these studies often made use of stimulus sets of (very) limited size, making generalizations towards the entire lexicon somewhat unfeasible. The current study employs a much larger dataset, comprising 4,151 unique words (contained in 665,461 word-pairs). With this, we were able to ascertain that the assortativity effects reported in existing research hold up for a large variability of cue stimuli.

Finally, we investigated word frequency, contextual diversity, and age-of-acquisition, three factors that are not related directly to the meaning of concepts, but rather reflect how a word is acquired by a speaker. Existing research reports that these aspects are all involved with connectivity in the mental lexicon: concepts that are learned at a young age show higher connectivity (Barabási & Albert, 1999; Steyvers & Tenenbaum, 2005), as are words with a high word frequency (Steyvers & Tenenbaum, 2005) and words with a high contextual diversity (Hills, Maouene, Riordan, & Smith, 2010). Note that while this indicates that these aspects are involved with the structure of the mental lexicon, we do not necessarily expect them to exhibit assortativity, which considers similarity between connected concepts and is distinct from overall connectivity. Indeed, our results show only a small cue-response correspondence for age-of-acquisition and virtually no correspondence on word frequency and contextual diversity, indicating that these aspects do not display assortativity in the mental lexicon.

From the previous discussion, it should be clear that assortativity describes how the mental lexicon is structured, but in itself does not directly inform us about causality. This raises the question whether factors that display assortativity actually influence response tendencies, or whether they simply co-vary with the type of responses made in an association task. In other words, do we produce a negative response to a negative cue because of their congruency in valence, or because they have similar (negative) meanings? It is often assumed that semantic similarity is the strongest determinant of response tendencies (Mollin, 2009), yet this does not necessarily rule out any influence of the psycho-affective properties of a word: these properties could correspond to semantic features, in which case the likelihood that the response depends on similarity to the cue would increase.

An alternative is to consider the word association process as reflecting learned co-occurrences derived from the linguistic environment. In this view, valence assortativity reflects negative or positive words co-occurring in language. The validity of this perspective could be addressed easily by examining assortativity in text corpora, and should be part of future investigations. However, we are very cautious at presenting this as a comprehensive explanation, as it has been pointed out on several occasions that by virtue of *not* being propositional, word associations capture different information than what can be inferred from a linguistic environment that conveys communicative constraints such as pragmatics (e.g., McRae, Khalkhali, & Hare, 2012; Szalay & Deese, 1978; De Deyne, Verheyen, Storms, 2015).

Assortativity effects have implications for studies in other domains, such as in research on priming. Firstly, assortativity as measured through word associations can be used to predict which factors will exhibit prime-target congruency effects, and which factors do not. For example, our findings are in line with the affective priming effect, where an affectively congruent prime facilitates processing more than an affectively incongruent prime (Fazio, 2001; Klauer, 1997; Spruyt, Hermans, De Houwer, Vandekerckhove, & Eelen, 2007). However, the current findings also point towards the fact that not all types of congruencies are equally strong, and that other factors can enhance or diminish these effects. For example, our findings suggest a larger congruency effect for valence than for concreteness; while these factors have been investigated separately in the priming literature, to our knowledge, they have not been compared directly. Moreover, our results also suggest strong effects for part of speech, which suggests that this factor should be controlled for when investigating congruency effects of other factors, such as in affective priming. Conversely, this relation between cue-target assortativity and congruency effects in priming research might also lead to new factors being included in future investigations of assortativity; for example, since a congruency effect of modality has been established in the priming literature (Pecher, Zeelenberg, & Barsalou, 2003), one might expect cue-target pairs to correspond on this dimension, too.

Common to all these cases is the idea that affectivity, modality, and concreteness might be part of a hierarchy of semantic properties, where valence is relevant to most words in the lexicon, while modality (visual, haptic) applies only to a subset of word, and specific semantic properties (e.g., “is an animal”) to even smaller regions of the lexicon.

In summary, the present research investigated the extent to which various word covariates exhibit assortativity in the mental lexicon. We find assortative mixing by valence, dominance, arousal, concreteness, and part of speech, but not by word frequency, contextual diversity, and age-of-acquisition.

References

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–512. doi:10.1126/science.286.5439.509

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 582–594. doi:10.1037//0278-7393.13.4.582

Bock, J. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, *150*, 80–84. doi:10.1016/j.actpsy.2014.04.010

Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407.

Collins, A., & Quillian, M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247.

Cramer, P. (1968). *Word association*. New York, NY: Academic Press.

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–98. doi:10.3758/s13428-012-0260-7

De Deyne, S., Navarro, D. J., & Storms, G. (2015). Associative strength and semantic activation in the mental lexicon: evidence from continued word associations. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*.

De Deyne, S., Verheyen, S., & Storms, G. (2015). The role of corpus size and syntax in deriving lexico-semantic representations for a wide range of concepts. *Quarterly Journal of Experimental Psychology*, *68*(8), 1643–1664. doi:10.1080/17470218.2014.994098

de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 824–845. doi:10.1037//0278-7393.15.5.824

Deese, J. (1962). Form class and the determinants of association. *Journal of Verbal Learning and Verbal Behavior*, *1*(2), 79–84. doi:10.1016/S0022-5371(62)80001-2

Deese, J. (1965). *The structure of associations in language and thought*. Baltimore, MD: Johns Hopkins University Press.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, *15*(2), 115–141. doi:10.1080/02699930125908

Grömping, U. (2006). Relative importance for linear regression in R: the package relaimpo. *Journal of Statistical Software*, *17*(1).

Heise, D. R. (2010). *Surveying cultures: Discovering shared conceptions and sentiments*. Hoboken, NJ: John Wiley & Sons.

Hills, T., Maouene, J., Riordan, B., & Smith, L. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273.

Hotopf, W. H. N. (1980). Semantic similarity as a factor in whole-word slips of the tongue. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 97–110). New York, NY: Academic Press.

Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2008). Predicting semantic priming at the item level. *Quarterly Journal of*

Experimental Psychology, *61*(7), 1036–1066. doi:10.1080/17470210701438111

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. New York, NY: Oxford University Press.

Johnson, R. C., & Lim, D. (1964). Personality variables in associative production. *Journal of General Psychology*, *71*(2), 349–350.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–50. doi:10.3758/BRM.42.3.643

Klauer, K. C. (1997). Affective Priming. *European Review of Social Psychology*, *8*(1), 67–103. doi:10.1080/14792779643000083

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. (2014). Emotion and Language: Valence and Arousal Affect Word Recognition. *Journal of Experimental Psychology. General*, *143*(3), 1065–1081. doi:10.1037/a0035669

Matlin, M. W., & Stang, D. J. (1978). *The Pollyanna principle: Selectivity in language, memory, and thought*. Cambridge, MA: Schenkman.

Mätzig, S., Druks, J., Masterson, J., & Vigliocco, G. (2009). Noun and verb differences in picture naming: Past studies and new evidence. *Cortex*, *45*, 738–758. doi:10.1016/j.cortex.2008.10.003

McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and associative relations in adolescents and young adults: Examining a tenuous dichotomy. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), *The adolescent brain: Learning, reasoning, and decision making* (pp. 39–66). Washington, DC: American Psychological Association. doi:10.1037/13493-002

Mollin, S. (2009). Combining corpus linguistic and psychological data on word co-occurrences: Corpus collocates versus word associations.

Corpus Linguistics and Linguistic Theory, *5*(2), 175–200. doi:10.1515/CLLT.2009.008

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A.-L., ... Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods*, *45*(1), 169–77. doi:10.3758/s13428-012-0243-8

Murphy, M. L. (2003). *Semantic relations and the lexicon*. Cambridge, U.K.: Cambridge University Press.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–7. doi:10.3758/BF03195588

Newman, M. (2010). *Networks: an introduction*. New York, NY: Oxford University Press.

Niedenthal, P., Halberstadt, J. B., & Innes-Ker, A. H. (1999). Emotional response categorization. *Psychological Review*, *106*(2), 336–361. doi:10.1037/0033-295X.106.2.337

Osgood, C. E. (1975). *Cross-cultural universals of affective meaning*. Urbana, IL: University of Illinois Press.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, *14*(2), 119–124.

Pickering, M., & Branigan, H. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, *3*(4), 136–141.

Pickering, M. J., & Branigan, H. P. (1998). The Representation of Verbs: Evidence from Syntactic Priming in Language Production. *Journal of*

Memory and Language, *39*(4), 633–651. doi:10.1006/jmla.1998.2592

Pollio, H. (1964). Some semantic relations among word-associates. *The American Journal of Psychology*, *77*(2), 249–256. doi:10.2307/1420131

Spruyt, A., Hermans, D., De Houwer, J., Vandekerckhove, J., & Eelen, P. (2007). On the predictive validity of indirect attitude measures: Prediction of consumer choice behavior on the basis of affective priming in the picture-picture naming task. *Journal of Experimental Social Psychology*, *43*, 599–610. doi:10.1016/j.jesp.2006.06.009

Staats, A. W., & Staats, C. K. (1959). Meaning and *m*: correlated but separate. *Psychological Review*, *66*(2), 136–44.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78. doi:10.1207/s15516709cog2901_3

Szalay, L. B., & Deese, J. (1978). *Subjective meaning and culture: An assessment through word associations*. Hillsdale, NJ: Lawrence Erlbaum.

Van Den Bosch, A., Busser, B., Canisius, S., & Daelemans, W. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch. *LOT Occasional Series*, *7*, 191–206.

Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research*, *51*(2), 408–422. doi:10.1044/1092-4388(2008/030)

Vitevitch, M. S., Chan, K. Y., & Goldstein, R. (2014). Insights into failed lexical retrieval from network science. *Cognitive Psychology*, *68*, 1–32. doi:10.1016/j.cogpsych.2013.10.002

Chapter 3

Estimating affective word covariates

This chapter was published as:

Van Rensbergen, B., De Deyne, S., & Storms, G. (2015). Estimating affective word covariates using word association data. *Behavior Research Methods*. Advance online publication.
doi: 10.3758/s13428-015-0680-2

Abstract

Word ratings on affective dimensions are an important tool in psycholinguistic research. Traditionally, they are obtained by asking participants to rate words on each dimension, a time-consuming procedure. As such, there has been some interest in computationally generating norms, by extrapolating words' affective ratings using their semantic similarity to words for which these values are already known. So far, most attempts have derived similarity from word co-occurrence in text corpora. In the current paper, we obtain similarity from word association data. We use these similarity ratings to predict the valence, arousal, and dominance of 14,000 Dutch words with the help of two extrapolation methods: Orientation towards Paradigm Words and *k*-Nearest Neighbors. The resulting estimates show very high correlations with human ratings when using Orientation towards Paradigm Words, and even higher correlations when using *k*-Nearest Neighbors. We discuss possible theoretical accounts of our results and compare our findings with previous attempts at computationally generating affective norms.

3.1 Introduction

Emotionally charged concepts are processed differently than emotionally neutral concepts. This intuitive idea is supported by research in multiple domains, including brain imaging (Lane, Chua, & Dolan, 1999; Lang et al., 1998; Maddock, Garrett, & Buonocore, 2003; Mourao-Miranda et al., 2003), semantic categorization (Moffat, Siakaluk, Sidhu, & Pexman, 2015; Newcombe, Campbell, Siakaluk, & Pexman, 2012; Niedenthal, Halberstadt, & Innes-Ker, 1999), affective priming (Fazio, 2001; Klauer, 1997), word associations (Cramer, 1968; Isen, Johnson, Mertz, & Robinson, 1985; Johnson & Lim, 1964; Matlin & Stang, 1978; Pollio, 1964), or word recognition reaction times (De Houwer, Crombez, Baeyens, & Hermans, 2001; Kuperman, Estes, Brysbaert, & Warriner, 2014).

Research on the emotional aspect of words traditionally makes use of three dimensions: (1) valence or evaluative attitude, generally rated on a good/bad or happy/unhappy scale, (2) arousal or activity, often represented on an active/passive scale, and (3) dominance or potency, usually expressed on a strong/weak or dominant/submissive scale. The importance of these dimensions was first described by Osgood, Suci, and Tannenbaum (1957). In an undertaking to quantify connotative meaning, they performed a factor analysis on a large number of verbal judgments of a wide variety of concepts and found that most of the variance in emotional assessments was accounted for by these three affective dimensions. Subsequent research has replicated these findings across dozens of cultures (see Heise, 2010, or Osgood, 1975, for an overview), indicating that the importance of these factors may be near universal.

Word values on these dimensions are commonly used both for investigating the influence of affective meaning on some other aspect, and to control for a possible confounding effect of the emotional charge of stimuli. As such, it is not surprising that there is a high demand for databases with affective norming data.

Traditionally, these norms are obtained by asking participants to rate a large number of words on each dimension. This procedure can be very expensive and time-consuming, as multiple persons have to rate each word in order to arrive at reliable measures (by means of average ratings). As a result, most norming databases are rather limited in the number of different words they contain, making generalizations towards the entire lexicon somewhat unfeasible. For example, the original Affective Norms for English Words (ANEW) dataset, likely the most frequently used norms, contains 'just' 1,034 unique words (Bradley & Lang, 1999). Despite the cumbersome nature of gathering ratings word by word, some researchers have recently managed to construct a much more comprehensive English database, containing norms for 13,915 words (Warriner, Kuperman, & Brysbaert, 2013). Affective rating datasets in other languages are not nearly as extensive, such as in Dutch (4,300 words: Moors et al., 2013), Finnish (420 words: Söderholm, Häyry, Laine, & Karrasch, 2013), French (1,031 words: Monnier & Syssau, 2014), German (2,900 words: Võ et al., 2009), Italian (1,034 words: Montefinese, Ambrosini, Fairfield, & Mammarella, 2014), Spanish (1,034 words: Redondo, Fraga, Padrón, & Comesaña, 2007), Polish (1,586 words: Imbir, 2015), or Portuguese (1,034 words: Soares, Comesaña, Pinheiro, Simões, & Frade, 2012).

3.1.1 Estimating affective ratings using word co-occurrence data

As the procedure of having participants rate words manually is both expensive and time-consuming, there has been some interest in deriving affective norms from other sources of information. One approach that has been suggested starts by deriving similarity measures for large numbers of words using their position in text corpora. For any given word in the corpus, norm ratings are then estimated using that word's similarity to a number of words for which affective values are already known. This approach could lead to norming datasets significantly larger than those gathered using manual ratings, as large text corpora are available in many languages.

Two implementations of this technique have been put forward. A first approach makes use of latent semantic analysis (LSA; Landauer & Dumais, 1997), which quantifies the degree to which words are associated based on the assumption that similar words occur in similar pieces of text. LSA starts from a word by context matrix, where each cell contains how frequently that word occurs in that chunk of text (e.g., sentence, paragraph, or document). To diminish the influence of highly frequent words, a weighting function is applied to this matrix. Subsequently the most important dimensions (usually 300) are extracted from this matrix using singular value decomposition, yielding a relatively low-dimensional approximation of the original matrix. The similarity between any two words is then defined as the cosine of the angle between their corresponding row vectors in this new matrix. As a result, LSA can estimate the similarity between two words that never occur together, but do co-occur in similar contexts.

A second approach to predict similarity from text corpora makes use of pointwise mutual information (PMI: Church & Hanks, 1990; see also Bullinaria & Levy, 2007; Manning & Schütze, 1999), which derives relatedness from direct word co-occurrence rather than co-occurrence in contexts. Specifically, the PMI of two words x and y is defined as

$$\mathrm{PMI}(x, y) = \log_2 \frac{P(x, y)}{P(x)P(y)},$$

where $P(x,y)$ refers to the frequency of x and y co-occurring in some context divided by the total number of tokens in the corpus, and $P(x)$ and $P(y)$ refer to the frequency of x and y, respectively divided by the total number of tokens. Compared with LSA, the most prominent advantage of PMI is scalability, as it can be applied to corpora far larger than LSA can handle. Additionally, it has been suggested that PMI may be more plausible as a model of semantic organization (Recchia & Jones, 2009).

Once pairwise similarity estimates have been derived by applying either LSA or PMI to text corpora, one can estimate words' values on various dimensions using their similarity towards words for which the values on those dimensions are already known.

Turney and Littman (2003) predicted the valence of words using their similarity to a small number of paradigm words, words commonly used to describe very low or very high levels of valence (e.g., *good*, *bad*). They compared the predictions of this approach with binary manual ratings (words rated positive or negative) for 3,596 English words, and report a correlation of .65 when using similarity derived from LSA (on a corpus comprising 10 million tokens), and between .61 (corpus containing 10 million tokens) and .83 (corpus containing 100 billion tokens) when using similarity derived from PMI.

Bestgen and Vincze (2012) employed a somewhat different approach. Rather than examine a word's relation to a small number of seed words, they looked at its similarity to all words for which norming data exist: they define the estimated rating of words as the average of its *k* nearest neighbors included in the norming data, with *k* ranging from 1 to 50. Nearest neighbors were obtained from similarity indices between 17,350 English words, which were calculated by applying LSA to a corpus comprising 12 million tokens. The valence, arousal, and dominance of each of these words was then estimated as the average rating of its *k* nearest neighbors which were included in the ANEW norms. Note that a given word was never considered as one of its nearest neighbors, that is, predictions were based on a leave-one-out approach. Comparing obtained estimates with the ANEW norms they find the highest accuracy at $k = 30$, with a correlation of .71 for valence, .56 for arousal, and .60 for dominance.

Recchia and Louwerse (2015) used a comparable approach, with a number of differences. They obtained nearest neighbors through similarity measures derived with PMI rather than LSA, which allowed them to make use of a much larger corpus containing 1.6 billion English words. They also tested a wider array of values for neighborhood parameter *k*, with *k* ranging from 2 to 500. Additionally, instead of following a leave-one-out approach, predictions were based on the ratings of one dataset while accuracy was assessed through correspondence to ratings of a second dataset. This revealed correlations of up to .74 for valence (at $k = 15$), up to .57 for arousal (at $k = 40$), and up to .62 for dominance (at $k = 60$).

Finally, Mandera, Keuleers, and Brysbaert (2015) evaluated how the performance of these computational approaches is influenced by the size of available norming data. To that end, the 13,915 words in the Warriner et al. (2013) norms were split into a training set and a test set, using different splits (e.g., 90%/10% or 50%/50%). Similarity indices between all words were obtained through applying LSA or PMI to a corpus comprising 385 million tokens. These were then used to predict the valence, arousal, and dominance of words, with neighborhood parameter *k* set to 30 (the optimal value described by Bestgen & Vincze, 2012). They find that accuracy is somewhat reliant on the size of available norms. For example, when working with PMI-based similarity, increasing the training sample (i.e., the ratings that can contribute to the estimates) from 10% of the Warriner norms to 90% raises the correlation between the test sample and the norm ratings from .61 to .72 for valence, from .37 to .51 for arousal, and from .51 to .61 for dominance. (They also investigated a number of other extrapolation methods, all of which showed a similar or lower accuracy.)

Taken together, these studies indicate that ratings extrapolated from word co-occurrence data show medium to high correlations with human judgments, highlighting the usefulness of this computational approach. Moreover, the size of norming databases constructed using this method is likely to keep expanding in the coming years, as even more word corpora become available. This is especially useful for languages other than English, where existing norming datasets are often quite limited in size.

3.1.2 Word associations as a source of similarity

As we have seen, existing research on computationally estimating norms generally makes use of similarity values derived from word co-occurrences in text corpora. An alternate approach to obtaining similarity ratings is using word association data. In a word association task, participants respond with the first word(s) that come to mind after reading a certain cue word. A key assumption in using word associations to investigate meaning is that the probability of producing a certain response to a cue is a measure of the associative strength between cue and response in the mental lexicon (Cramer, 1968; De Deyne, Navarro, & Storms, 2013; Deese, 1965; Nelson, McEvoy, & Schreiber, 2004). This idea is supported by research on facilitation of word processing in associative priming (Hutchison, 2003), response times in lexical decision tasks (De Deyne et al., 2013), word recognition reaction times (De Deyne et al., 2013; Gallagher & Palermo, 1967; Nelson, McKinney, Gee, & Janczura, 1998), fluency task generation frequencies (Griffiths, Steyvers, & Tenenbaum, 2007), clustering in recall (Wicklund, Palermo, & Jenkins, 1965), and predicting cued recall (Nelson et al., 1998).

To obtain information about relatedness from word association data, one can make use of a cosine measure of similarity (Landauer & Dumais, 1997). While this measure is traditionally applied to spatial models such as LSA, it can also be used in the context of word association data (e.g., De Deyne et al., 2013; De Deyne, Verheyen, & Storms, 2015; Gravino, Servedio, Barrat, & Loreto, 2012). Here, the cosine similarity between two words reflects their overlap in associative links; two words that share no associations have a similarity of 0, while two words with the exact same associative responses have a similarity of 1. Similarity estimates obtained using this approach show a strong correspondence with relatedness judgments (De Deyne et al., 2013; De Deyne et al., 2015).

Research indicates that, compared with approaches based on text corpora, word association data can lead to a more valid measure of semantic relatedness. For example, (human) similarity judgments correlate more strongly with similarity estimates derived from association data than with predictions based on word co-occurrences (De Deyne, Peirsman, & Storms, 2009; De Deyne et al., 2015). Additionally, associative strength has been shown to predict priming effects on a word-level in both lexical decision tasks and naming tasks, while similarity derived from applying LSA to text corpora did not (Hutchison, Balota, Cortese, & Watson, 2008).

In the current study, we propose using word association data to obtain similarity estimates for a large number of words, and subsequently predict words' values on affective dimensions (e.g., valence) using their similarity towards words for which the values on those dimensions are already known (e.g., *pleasant*). Using this approach, we will estimate valence, arousal, and dominance ratings for a large number of words. To verify the validity of these estimates, we will compare them with existing norm ratings.

3.2 Method

3.2.1 Materials

To obtain the associative strength for a large set of words, we made use of the Dutch Small World of Words project, which contains 3.7 million word associations collected in response to 14,000 cue words. Each cue was presented to roughly 100 participants, who gave up to three responses per cue (see De Deyne et al., 2013, for full details[4]).

Valence, arousal, and dominance ratings for 4,300 Dutch words were taken from Moors et al. (2013). In this study, words were rated on a Likert scale ranging from 1 (*very negative/unpleasant*, *very passive/calm*, and *very weak/submissive*, respectively) to 7 (*very positive/pleasant*, *very active/aroused*, and *very strong/dominant*). Ratings showed very high split-half reliabilities: .99 for valence, .97 for arousal, and .96 for dominance.

[4] We use a more recent version of this dataset, which is somewhat larger (e.g., comprising 14,000 cue words rather than 12,000) but otherwise similar in all aspects.

3.2.2 Procedure

We began by computing the cosine similarity (e.g., Landauer & Dumais, 1997) between each combination of the 14,000 cue words in the Dutch Small World of Words dataset. In this context, a cosine measure reflects the extent to which two words overlap in associative responses: two words that share no associations would have a value of 0, while two words with the exact same associative responses would have a value of 1. To obtain this measure, we first constructed a cue by cue count matrix, where cells reflected how often each cue was given as an association in response to each other cue. Rows of this matrix where normalized to sum to 1 and log-transformed. Finally, to obtain the cosines between the angles of these vectors, the matrix was multiplied by its transpose. At this point, cells of the matrix contained the cosine similarity between the cues corresponding to their rows and columns.

Subsequently, we used these similarity ratings to predict affective word covariates by applying two extrapolation methods, each of which estimates word's values on affective dimensions using that word's similarity to certain words for which affective ratings are already known.

The first extrapolation method we employed, *Orientation towards Paradigm Words*, predicted a word's valence, arousal, and dominance using that word's similarity towards certain paradigm words, words commonly used to describe extreme values on these dimensions (Kamps, Marx, Mokken, & de Rijke, 2004; Turney & Littman, 2003). Paradigm words were obtained from the instructions in the rating task described by Moors et al. (2013), which yielded two positive and two negative paradigm words for each dimension (Table 3.1).

At first, Orientation towards Paradigm Words predictions simply reflected the sum of a word's similarity towards both positive paradigm words minus the sum of its similarity towards both negative paradigm words. These estimates were consequentially refined by including the target word's similarities towards the *k* nearest neighbors of each of the paradigm words, that is, out of the 14,000 words, the *k* words with the highest similarity towards that paradigm word, where *k* ranged from 0 to 500. A target word's final score was computed as the sum of its similarity towards both positive paradigm words and the *k* nearest neighbors of each positive paradigm word, minus the sum of its similarity towards both negative paradigm words and the *k* nearest neighbors of each negative paradigm word.

The second extrapolation method we applied, k-*Nearest Neighbors*, was very similar to the approach described by Bestgen and Vincze (2012), with the notable difference that our similarity estimates were derived from word association data rather than from word co-occurrence in text corpora. Under this approach, the score of any target word on some dimension is calculated as the mean score of its *k* nearest neighbors (as assessed with cosine similarity) for which the value on that dimension is known (that is, the *k* closest words for which human judgments are included in the dataset of Moors et al., 2013), for *k* ranging from 1 to 500. Note that a target word is never considered as one of its own nearest neighbors; as such, the human judgment of some word does not contribute to that word's extrapolated rating.

It may be important to stress that with the *k*-Nearest Neighbors approach, *k* refers to the nearest neighbors of the target word (for which ratings were available), while under the Orientation towards Paradigm Words method, *k* refers to the nearest neighbors of the various paradigm words.

Table 3.1

English translation of the paradigm words corresponding to valence, arousal, and dominance (Dutch source that was actually used)

Dimension	Positive Paradigm Words (*Dutch source*)	Negative Paradigm Words (*Dutch source*)
Valence	positive, pleasant (*positief, aangenaam*)	negative, unpleasant (*negatief, slecht*)
Arousal	Active, busy (*actief, druk*)	passive, calm (*passief , kalm*)
Dominance	strong, dominant (*sterk, dominant*)	weak, submissive (*zwak, onderdanig*)

3.3 Results

We estimated the valence, arousal, and dominance of the 14,000 cue words in the Small World of Words dataset with the two extrapolation methods described above. Out of these 14,000 words, 3,872 are comprised in the norms of Moors et al. (2013) and can be used to assess the accuracy of the two methods. These 3,872 words represent 90% of the 4,300 words in the norms, and 28% of the cue words in the word association dataset.

The Orientation towards Paradigm Words method predicted the affective values of words using their similarity towards certain paradigm words (see Table 3.1), and the *k* nearest neighbors of each paradigm word. The left panel of Table 3.2 displays the correlations (Pearson's *r*) between these estimates and the human judgments described by Moors and colleagues (2013) for valence, arousal, and dominance, for *k* values ranging from 0 (only the paradigm words themselves are used) to 500 (the paradigm words and the 500 nearest neighbors of each paradigm word contribute to the final estimate).[5] When estimates are based solely on similarity to the paradigm words themselves, we find correlations of .79, .53, and .59 to human judgments of valence, arousal, and dominance, respectively. As we increase the number of neighbors of each paradigm word that contribute to our predictions, these correlations increase to up to .86, .65, and .69 for valence, arousal, and dominance, respectively.

---

[5] We also investigated the effect of applying various monotonically decreasing weighting functions to the contribution of the various nearest neighbors of each paradigm word, so the similarity towards further neighbors contributed less to the final score. Somewhat contrary to our expectations, none of these functions led to a significant improvement in the overall accuracy of our approach; as such, these findings are not reported here.

The *k*-Nearest Neighbors method estimated the valence, arousal, and dominance of the 14,000 words as the mean of the human ratings of its *k* nearest neighbors included in the Moors et al. (2013) dataset. The right panel of Table 3.2 displays the Pearson correlation between these estimates and human judgments of valence, arousal, and dominance, for *k* (the number of neighbors of a target word that contribute to its estimate) ranging from 1 to 500.[6] We find an optimal accuracy at $k = 10$, where the extrapolated ratings show a correlation of .91 for valence, .84 for arousal, and .85 for dominance.

We find that performance of both extrapolation methods shows a curvilinear function with respect to neighborhood parameter *k*: as *k* increases, accuracy improves up to a certain point and then starts to decline. This decreased performance at higher values of *k* is in line both with expectations, as 'further' neighbors have a lower similarity to the target word, and with previous research (Recchia & Louwerse, 2015).

A downside of the *k*-Nearest Neighbors approach is that it relies on an existing set of human judgments. As a result, the number of words for which human ratings are available is certain to have an effect on the accuracy of this method. If only few norms are available, it is possible that some extrapolated values are based on ratings of words that are in fact not particularly close to the target word (if more similar words are not included in the norming dataset), which would certainly have consequences for the validity of those estimates. In Dutch, we have access to 3,872 words in the relatively large norms of Moors et al. (2013); in many languages, databases of this size are not available. To estimate how

[6] Here, too, we examined the effect of applying different weighting functions to these k values, with further neighbors contributing less to a target word's final score. As with the first extrapolation method, this did not lead to a considerable improvement in accuracy; as such, we will not report these findings here.

accurate this extrapolation method would be when only a limited set of norms is available, we followed an approach similar to that of Mandera et al. (2015) by running the *k*-Nearest Neighbors method restricted to random subsets of the available norming data (at $k = 10$, the optimal value in Table 3.1). We tested 12 different sample sizes, ranging from 100 words to 3,872 words (the entire dataset). To remove any sampling bias, this procedure was repeated 100 times for each sample size. Figure 3.1 indicates that even when only a small norming dataset is available, the *k*-Nearest Neighbors method manages to attain a high accuracy; for exampling, when norms for just 1,000 words are available, the extrapolated ratings show correlations with human judgments of up to .89 for valence, and up to .79 for arousal and dominance.

Finally, we wanted to have an idea of whether having access to a norming dataset larger than that of Moors et al. (2013) would lead to a significant improvement in accuracy. Although we cannot test this notion directly with the data currently at our disposal, we can estimate it by examining the slopes of the lines in Figure 3.1. As all three lines keep increasing up to the largest sample size, it seems reasonable to assume that expanding the size of the used norming dataset would result in a small improvement in accuracy, especially for arousal and dominance.

Table 3.2

Correlations between human judgments and estimates derived using the Orientation towards Paradigm Words extrapolation method (left panel) and estimates derived using the k*-Nearest Neighbors extrapolation method (right panel)*

	Orientation towards Paradigm Words			*k*-Nearest Neighbors		
k	Val.	Ar.	Dom.	Val.	Ar.	Dom.
0	.79	.53	.59	-	-	-
1	.80	.54	.60	.85	.76	.76
2	.80	.54	.62	.88	.80	.81
5	.79	.49	.63	.89	.83	.83
10	.81	.56	.67	.91	.84	.85
25	.83	.63	.69	.91	.84	.84
50	.84	.63	.69	.91	.83	.83
100	.84	.67	.68	.91	.83	.82
250	.85	.65	.68	.90	.81	.81
500	.86	.63	.68	.90	.78	.79

Note. n = 3,872. Val. = Valence, Ar. = Arousal, Dom. = Dominance.

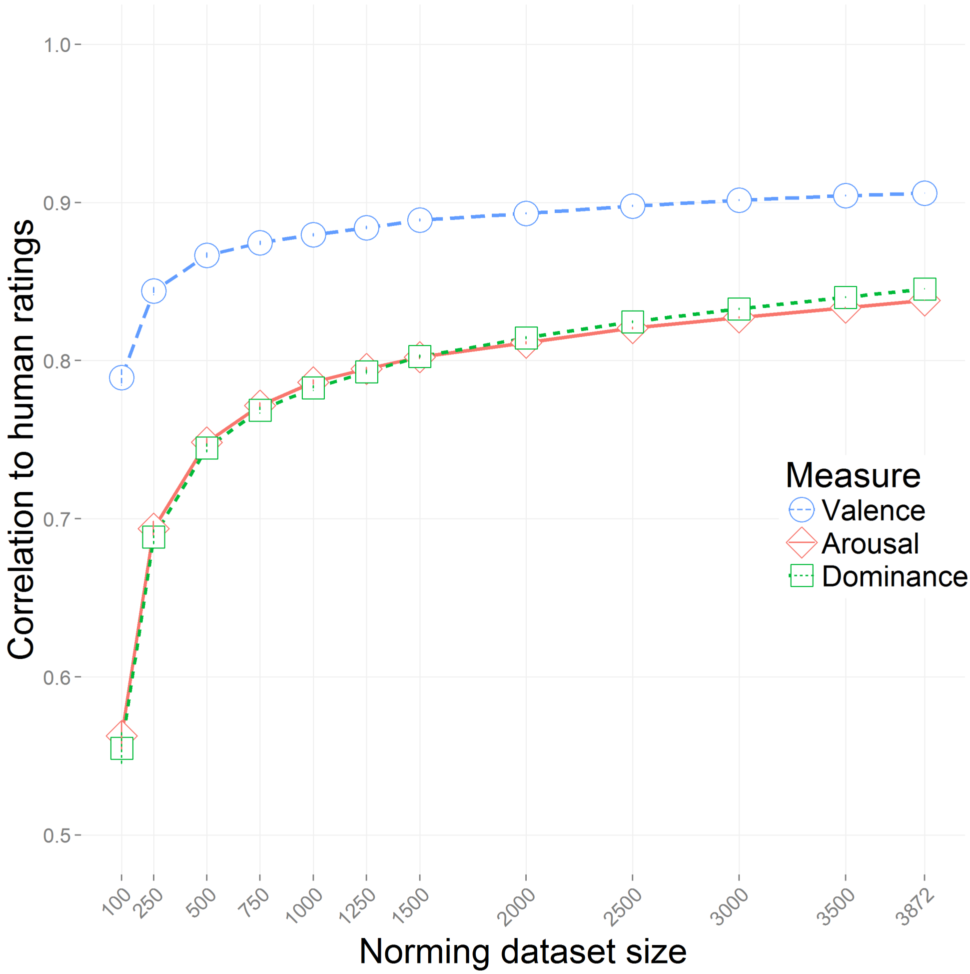

Figure 3.1. Relation between accuracy of the *k*-Nearest Neighbors extrapolation method and the size of available norms. Correlations were obtained by averaging across 100 iterations of running the extrapolation method limited to a random subset of human judgments (out of the available 3,872 norm words). Neighborhood parameter *k* was set to 10, the optimal value reported for running this extrapolation method with all human judgments (Table 3.2). Error bars (very small, due to low error rates) indicate standard error in accuracy across the 100 iterations.

3.4 Discussion

We have outlined two methods to computationally estimate subjective norms values. Both methods derive similarity from association data, and predict a word's norms using its similarity towards words for which affective values are already known. The two approaches were used to extrapolate the valence, arousal, and dominance for 14,000 Dutch words; these estimates are available at `https://osf.io/pmbvc/`.

In comparing the extrapolated norms to human judgments, we find high to very high correlations for all three dimensions. Correspondence is highest for valence, suggesting that compared with arousal and dominance, valence is represented more strongly in the semantic similarity space. This finding is in line with the importance often attributed to this aspect, both in research on affective meaning (Osgood et al., 1957) and various other domains.

Of the two extrapolation methods we tested, accuracy is highest for the *k*-Nearest Neighbors technique, as would be expected because this method is based directly on the human ratings with which accuracy is assessed (although importantly, the human judgment of a given word does not contribute to that word's extrapolated value). Note, though, that this reliance on human ratings brings with it a huge drawback: the *k*-Nearest Neighbors method can only work when human judgments are already available for some amount of words. In contrast, the Orientation towards Paradigm Words approach does not depend on human judgments in any form, and aside from a selection of paradigm words, only requires similarity indices.

Considering the *k*-Nearest Neighbors method relies on human judgments, its accuracy is likely tied to the quality of available human ratings. As our research was performed in Dutch, we had access to the large norming dataset of Moors and colleagues (2013). In many languages, existing databases are considerably smaller. To assess how accurate our approach is when limited to a smaller set of norms, we ran the *k*-Nearest Neighbors extrapolation method restricted to subsets of the available norming data. Correlations with human judgments were lower than when the method had access to all norming data, but still very high (between .78 and .88 when using a subset of 1,000 words). This suggests that even when only a small set of norms is available, the *k*-Nearest Neighbors method can be very effective at predicting affective word covariates.

In existing research on computationally predicting affective norms, similarity or semantic relatedness is generally derived from word co-occurrence data rather than from word associations. Using these similarity estimates, several studies have extrapolated affective ratings with the help of the same *k*-Nearest Neighbors technique we described. These studies report that their estimates display correlations with human judgments of up to .74, .57, and .62 (Bestgen & Vincze, 2012), up to .71, .56, and .60 (Recchia & Louwerse, 2015), and roughly up to .72, .51, and .61 (Mandera et al., 2015), for valence, arousal, and dominance, respectively.

In comparison, the predictions we present show a much higher accuracy, on all three dimensions. There are several potential explanations behind this improvement. It could be a result of a difference in language: we made use of Dutch associations and judgments, while the described corpus-based studies were performed in English. However, this seems an unlikely explanation, as similar corpus-based research has also been undertaken in French and Spanish, where estimates showed similar or lower correlations with human ratings (Bestgen, 2002, 2008; Vincze & Bestgen, 2011). Furthermore, as the importance of valence, arousal, and dominance is highly generalizable across cultures (Osgood, 1975), there is no a priori reason to expect these aspects to be represented differently in Dutch and English.

A more probable cause for the disparity between our findings and previous attempts at computationally estimating norms is the nature of the information from which similarity estimates were construed: existing research derived relatedness from word co-occurrence in text corpora, while we made use of word association data. Previous comparisons between corpus-based and association-based similarity estimates also report a higher accuracy for approaches reliant on word association data, in line with our findings (De Deyne et al., 2009; De Deyne et al., 2015; Hutchison et al., 2008). This is likely because word associations and text corpora represent information of a different nature. Written language is grounded in pragmatics; the goal is to communicate some discourse efficiently, and information that is known to both parties is often left out. Word associations, in contrast, are non-propositional, and generally free from pragmatics or intent (Deese, 1965; Szalay & Deese, 1978). As a result, mentally central concepts or properties (such as color or shape) are usually well represented in word associations, while they are somewhat uncommon in most written text. An additional asset of word association data is its very high signal to noise ratio, as almost every association reflects a meaningful relation; in contrast, text corpora are often characterized by a low signal-to-noise ratio, negating part of the advantage of scale that characterizes corpus-based approaches.

Taken together, we can conclude that word association data can be a very powerful source of information on semantic relatedness, and suggest that when computationally generating affective norms, an association-based approach may be a worthwhile addition to or substitute for procedures based on word co-occurrence in text corpora.

Of course, this approach does require access to word association data. While gathering word associations is a simple and straightforward procedure, it remains reliant on human participants. As a result, constructing a large dataset of this nature is far from effortless. Luckily, such databases already exist in many languages; for example, the Small World of Words project from which we obtained the Dutch associations also contains datasets in English, German, French, Spanish, Rioplatense Spanish, Vietnamese, Japanese, and Cantonese. Note that in terms of number of tokens, these databases are all much smaller than most text corpora. However, as we have seen, this quantitative shortcoming does not necessarily translate to deteriorated predictions; indeed, human judgments show a considerably higher correspondence to the estimates reported in the current paper, which are based on a dataset comprising 3.7 million tokens, than to the estimates based on word co-occurrence data described previously, which are based on much larger corpora (e.g., the predictions reported by Recchia & Louwerse, 2015, are based on a dataset containing 1.6 billion tokens).

An important caveat when working with computationally estimated word covariates is that even when they show a moderate to high correspondence with human judgments, they could lead to different conclusions than would be reached when using human ratings (Mandera et al., 2015). The data we present are likely somewhat less vulnerable to this issue, as our estimates show considerably higher correlations to human ratings; nevertheless, this is definitely a topic that should be investigated further in future research.

In the current paper, we estimated valence, arousal, and dominance ratings based on similarity values derived from word association data. The extrapolation methods we describe would conceivably work on other psychologically relevant dimensions as well, as long as these dimensions are captured by the associative technique, that is, as long as the associations people give to a certain word are in some way related to the cue's or association's value on that dimension. Existing research suggests that other examples of word covariates that could likely be predicted based on association data may include concreteness (the extent to which words refer to something perceptible; see Mandera et al., 2015, or Van Rensbergen, Storms, & De Deyne, 2015), age of acquisition (the age at which a word was learned; see Mandera et al., 2015) or dimensions relevant to personality profiles (e.g., openness, conscientiousness, extraversion, agreeableness, or neuroticism; see Yarkoni, 2010, or Park et al., 2015).

References

Bestgen, Y. (2002). Détermination de la valence affective de termes dans de grands corpus de textes. *Actes Du Colloque International Sur La Fouille de Texte*, *2*, 81–94.

Bestgen, Y. (2008). Building Affective Lexicons from Specific Corpora for Automatic Sentiment Analysis. *Proceedings of LREC*, 496–500.

Bestgen, Y., & Vincze, N. (2012). Checking and bootstrapping lexical norms by means of word similarity indexes. *Behavior Research Methods*, *44*(4), 998–1006. doi:10.3758/s13428-012-0195-z

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Gainesville, FL: University of Florida.

Bullinaria, J., & Levy, J. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526. doi:10.3758/BF03193020

Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29.

Cramer, P. (1968). *Word association*. New York, NY: Academic Press.

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–98. doi:10.3758/s13428-012-0260-7

De Deyne, S., Peirsman, Y., & Storms, G. (2009). Sources of semantic similarity. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (pp. 1834–1839). Austin, TX: Cognitive Science Society.

De Deyne, S., Verheyen, S., & Storms, G. (2015). The role of corpus size and syntax in deriving lexico-semantic representations for a wide range of concepts. *Quarterly Journal of Experimental Psychology*, *68*(8), 1643–1664. doi:10.1080/17470218.2014.994098

De Houwer, J., Crombez, G., Baeyens, F., & Hermans, D. (2001). On the generality of the affective Simon effect. *Cognition & Emotion*, *15*(2), 189–206. doi:10.1080/02699930125883

Deese, J. (1965). *The structure of associations in language and thought*. Baltimore, MD: Johns Hopkins University Press.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, *15*(2), 115–141. doi:10.1080/02699930125908

Gallagher, J., & Palermo, D. (1967). The effect of type associative relation on word recognition times. *Child Development*, *38*(3), 849–855. doi:10.2307/1127262

Gravino, P., Servedio, V. D. P., Barrat, A., & Loreto, V. (2012). Complex structures and semantics in free word association. *Advances in Complex Systems*, *15*(03n04). doi:10.1142/S0219525912500543

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244. doi:10.1037/0033-295X.114.2.211

Heise, D. R. (2010). *Surveying cultures: Discovering shared conceptions and sentiments*. Hoboken, NJ: John Wiley & Sons.

Hutchison, K. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813. doi:10.3758/BF03196544

Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2008). Predicting semantic priming at the item level. *Quarterly Journal of*

Experimental Psychology, *61*(7), 1036–1066. doi:10.1080/17470210701438111

Imbir, K. K. (2015). Affective norms for 1,586 polish words (ANPW): Duality-of-mind approach. *Behavior Research Methods*, *47*(3), 860-870. doi:10.3758/s13428-014-0509-4

Isen, A. M., Johnson, M. M., Mertz, E., & Robinson, G. F. (1985). The influence of positive affect on the unusualness of word associations. *Journal of Personality and Social Psychology*, *48*(6), 1413–26. doi:10.1037/0022-3514.48.6.1413

Johnson, R. C., & Lim, D. (1964). Personality variables in associative production. *Journal of General Psychology*, *71*(2), 349–350.

Klauer, K. C. (1997). Affective Priming. *European Review of Social Psychology*, *8*(1), 67–103. doi:10.1080/14792779643000083

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. (2014). Emotion and Language: Valence and Arousal Affect Word Recognition. *Journal of Experimental Psychology. General*, *143*(3), 1065–1081. doi:10.1037/a0035669

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. doi:10.1037/0033-295X.104.2.211

Lane, R. D., Chua, P. M. L., & Dolan, R. J. (1999). Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia*, *37*, 989–997. doi:10.1016/S0028-3932(99)00017-2

Lang, P. J., Bradley, M. M., Fitzsimmons, J. R., Cuthbert, B. N., Scott, J. D., Moulder, B., & Nangia, V. (1998). Emotional arousal and activation of the visual cortex: an fMRI analysis. *Psychophysiology*, *35*, 199–210. doi:10.1111/1469-8986.3520199

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human Brain Mapping*, *18*, 30–41. doi:10.1002/hbm.10075

Mandera, P., Keuleers, E., & Brysbaert, M. (2015). How useful are corpus-based methods for extrapolating psycholinguistic variables? *The Quarterly Journal of Experimental Psychology*, *68*(8), 1623–1642. doi:10.1080/17470218.2014.988735

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT press.

Matlin, M. W., & Stang, D. J. (1978). *The Pollyanna principle: Selectivity in language, memory, and thought*. Cambridge, MA: Schenkman.

Moffat, M., Siakaluk, P. D., Sidhu, D. M., & Pexman, P. M. (2015). Situated conceptualization and semantic processing: effects of emotional experience and context availability in semantic categorization and naming tasks. *Psychonomic Bulletin & Review*, *22*(2), 408–419. doi:10.3758/s13423-014-0696-0

Monnier, C., & Syssau, A. (2014). Affective norms for French words (FAN). *Behavior Research Methods*, *46*(4), 1128–1137. doi:10.3758/s13428-013-0431-1

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the Affective Norms for English Words (ANEW) for Italian. *Behavior Research Methods*, *46*(3), 887–903. doi:10.3758/s13428-013-0405-3

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A.-L., ... Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods*, *45*(1), 169–77. doi:10.3758/s13428-012-0243-8

Mourão-Miranda, J., Volchan, E., Moll, J., De Oliveira-Souza, R., Oliveira, L., Bramati, I., ... Pessoa, L. (2003). Contributions of stimulus valence and arousal to visual activation during emotional perception. *NeuroImage, 20,* 1955–1963. doi:10.1016/j.neuroimage.2003.08.011

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–7. doi:10.3758/BF03195588

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, *105*(2), 299–324. doi:10.1037/0033-295X.105.2.299

Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of Emotional and Sensorimotor Knowledge in Semantic Processing of Concrete and Abstract Nouns. *Frontiers in Human Neuroscience*, *6*. doi:10.3389/fnhum.2012.00275

Niedenthal, P., Halberstadt, J. B., & Innes-Ker, A. H. (1999). Emotional response categorization. *Psychological Review*, *106*(2), 336–361. doi:10.1037/0033-295X.106.2.337

Osgood, C. E. (1975). *Cross-cultural universals of affective meaning*. Urbana, IL: University of Illinois Press.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... Seligman, M. E. P. (2015). Automatic Personality Assessment Through Social Media Language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. doi:10.1037/pspp0000020

Pollio, H. (1964). Some semantic relations among word-associates. *The American Journal of Psychology*, *77*(2), 249–256. doi:10.2307/1420131

Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*, *41*(3), 647–656. doi:10.3758/BRM.41.3.647

Recchia, G., & Louwerse, M. M. (2015). Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *Quarterly Journal of Experimental Psychology*, *68*(8), 1584–1598. doi:10.1080/17470218.2014.941296

Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW (affective norms for English words). *Behavior Research Methods*, *39*(3), 600–605. doi:10.3758/BF03193031

Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., & Frade, C. S. (2012). The adaptation of the Affective Norms for English Words (ANEW) for European Portuguese. *Behavior Research Methods*, *44*(1), 256–69. doi:10.3758/s13428-011-0131-7

Söderholm, C., Häyry, E., Laine, M., & Karrasch, M. (2013). Valence and arousal ratings for 420 Finnish nouns by age and gender. *PloS One*, *8*(8), e72859. doi:10.1371/journal.pone.0072859

Szalay, L. B., & Deese, J. (1978). *Subjective meaning and culture: An assessment through word associations*. Hillsdale, NJ: Lawrence Erlbaum.

Turney, P., & Littman, M. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, *21*(4), 315–346. doi:10.1145/944012.944013

Van Rensbergen, B., Storms, G., & De Deyne, S. (2015). Examining assortativity in the mental lexicon: Evidence from word associations. *Psychonomic Bulletin & Review, 22*(6), 1717-1724. doi:10.3758/s13423-015-0832-5

Vincze, N., & Bestgen, Y. (2011). Une procédure automatique pour étendre des normes lexicales par l'analyse des cooccurrences dans des textes. *Traitement Automatique Des Langues*, *52*(3), 191–216.

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–8. doi:10.3758/BRM.41.2.534

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–207. doi:10.3758/s13428-012-0314-x

Wicklund, D., Palermo, D., & Jenkins, J. (1965). Associative clustering in the recall of children as a function of verbal association strength. *Journal of Experimental Child Psychology*, *2*(1), 58–66. doi:10.1016/0022-0965(65)90015-9

Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373. doi:10.1016/j.jrp.2010.04.001

Chapter 4

Computationally coding a free-format personality test

This chapter was submitted as:

Van Rensbergen, B., Kuppens, P., Storms, G., & De Deyne, S. (2015). Computationally coding responses of a free-format self-description personality test using word association data. Manuscript submitted for publication.

Abstract

Objective: Deriving personality from free self-descriptions offers multiple advantages over fixed-format questionnaires and holds wide application potential. Yet it struggles with one large drawback: obtaining trait scores from responses. Traditionally, trained psychologists do this by rating correspondence of the descriptions to various personality domains, a rather extensive procedure. To resolve this, we propose a computational approach of encoding free-format responses.

Method: 71 participants completed a free-response personality test, in which they described their personality using any ten adjectives. Participants' personality scores were obtained from the average correspondence of their responses to the Big Five personality factors, which was encoded both by trained psychologists and computationally, using the responses' semantic similarity to words for which trait correspondence was already known. In a secondary analysis, obtained personality profiles were compared with output of the NEO-PI-3, a fixed-format questionnaire.

Results: Coding responses computationally yields results similar to ratings of trained psychologists, with a mean correlation of .84. Obtained scores show a correlation of .41 with output of the NEO-PI-3.

Conclusions: The proposed computational method of deriving personality from free-format self-descriptions is a viable approach. Advantages and potential applications of this method are discussed, as are theoretical accounts of our findings.

4.1 Introduction

Personality plays an important role in various aspects of life, including personal, social, and professional success, life satisfaction, health and sickness, tendency towards criminal behavior, and even life length (e.g., Pervin & John, 1999). As such, it is of vital importance to have access to fast and accurate measures of individual differences on personality domains.

A common approach to represent personality makes use of the Big Five structure, which encompasses the domains openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (Costa & McCrae, 1985, 1995; Norman, 1963; Tupes & Christal, 1961; see also Goldberg, 1993). Traditionally, the Big Five are measured using a fixed-format questionnaire, where participants answer a series of items on some response scale. For example, in the NEO-PI-3 (McCrae, Costa, & Martin, 2005) participants respond to 240 questions (e.g., "I sometimes act thoughtlessly") on a five-point scale ranging from *strongly disagree* to *strongly agree*.

An alternative approach to measuring personality is to derive it from word use. Research has indicated that the language that people use to express themselves corresponds to self-reports of personality; for example, personality has been linked to word use in daily diaries, writing assignments, and abstracts of peer-reviewed articles (Pennebaker & King, 1999), in online blogs (Yarkoni, 2010; Iacobelli, Gill, Nowson, & Oberlander, 2011), in Facebook status updates (Park et al., 2014), or in Twitter messages (Golbeck, Robles, Edmondson, & Turner, 2011). Importantly, it has been found that social media profiles reflect actual personality, rather than an idealized version of it (Back et al., 2010).

This connection between language and personality has also been used to develop free-format personality tests such as the free-response self-description method, in which participants are asked to describe their own personality using any 10 adjectives (Claeys, De Boeck, Van Den Bosch, Biesmans, & Böhrer, 1985; see also Potkay & Allen, 1973, for an overview of the test on which this free-response method is based). Obtained responses are scored on their correspondence to the Big Five by trained psychologists, and participants' profiles are created from the average ratings of their responses. Compared with fixed-format questionnaires, personality profiles obtained with this free-response assessment show moderate to high correlations, and display similar to higher validity (Claeys et al., 1985; S.P.O.L.A.P., 1987; Van Den Broucke, De Soete, & Böhrer, 1989). Additionally, the self-description test displays moderately high test-retest reliability, and very high scoring reliability (S.P.O.L.A.P., 1987).

Compared with fixed-format personality inventories, a free-format approach has the advantage of being a more immediate and natural task. It minimizes most interference by the researcher, such as a selection bias in terms of wording or item choice. Indeed, participants are not constrained at all in their responses, as they may give any answer they consider meaningful and are not forced to respond to questions they are unsure of or do not deem relevant to their personality (Claeys et al, 1985; Pervin, 1976). This is especially useful when conducting cross-cultural personality research, as items selected by the investigator may be irrelevant or hold a different meaning in some cultures (Poortinga, 1989), and participants of some cultures may be less familiar with rating or attitude scales (Peng, Nisbett, & Wong, 1997).

A second advantage of free-format personality tests comes to light when they are used in addition to fixed-format questionnaires, rather than replacing them altogether. Claeys and colleagues (1985) found that when their free-format self-description task preceded a questionnaire, the validity of the latter was raised significantly, perhaps because the former activated relevant self-knowledge by making people think consciously about their personality.

A practical advantage of free-format tests is that they take considerably less time per participant, who gives 10 adjectives as opposed to responding to a large number of items on a Likert scale.

Of course, free-format personality measures have some disadvantages as well. First, it is possible that rather than describe their personality to the best of their ability, participants respond with traits they would like to possess or believe to be socially desirable. Although fixed-format tests share this issue to some extent, they can include control questions, and it may be somewhat harder for participants to judge which response options would be the most socially acceptable. It is not obvious how a similar control mechanism could be built into a free-format measurement.

Second, while the free-response test may have no bias in terms of choice of items or descriptions, some degree of subjectivity likely plays a role when experts rate obtained responses on their correspondence to the various personality domains.

This brings us to perhaps the most prominent drawback of free-format assessments: the issue of deriving traits scores from obtained responses. For fixed-format questionnaires, this can be done computationally, an all but trivial enterprise. In contrast, free-format assessments have to be scored by multiple experts, who have to rate the correspondence of each response to various personality domains, a rather extensive and time-consuming procedure.

To facilitate this process, it is possible to code obtained responses using norm datasets that contain personality ratings of words (e.g., Anderson, 1968; Schönbach, 1972; S.P.O.L.A.P., 1987). However, datasets of this type exist in few languages and often contain only a very limited number of words. Creating a set of norms sufficiently large to properly encode all free-format responses would be a huge undertaking. As a result, a computational approach to encoding free-format response might prove useful.

In this paper, we will outline a technique to predict the correspondence of various words to the Big Five factors. The process begins by deriving semantic similarity indices from some word association data, and subsequently predicts words' values on various dimensions using that word's similarity to words for which these values are already known. We can then use these ratings to code responses obtained in a free-format personality assessment computationally.

Word associations are collected by asking participants to write down the first word(s) that come to mind spontaneously after reading a certain cue word. The probability that a cue elicits a certain response is considered a measure of the associative strength between that cue and response in the mental lexicon (De Deyne, Navarro, & Storms, 2015; Nelson, McEvoy, & Schreiber, 2004). This direct connection to what goes on in the mind makes word association data extremely useful when investigating subjective meaning (Deese, 1965) as it tends to be less censored (Szalay & Deese, 1978) and provides a more direct pathway to meaning compared with written or spoken text sources (McRae, Khalklhali & Hare, 2011; Mollin, 2009).

In a recent study, we predicted the valence, arousal, and dominance of words using similarity indices derived from word associations (Van Rensbergen, De Deyne, & Storms, 2015). This process proved to work rather well, as obtained estimates showed correlations with human ratings of up to .91 for valence and up to .84 for arousal and dominance.

In the current paper, we propose following the same approach to predict the correspondence of various words to the domains openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. We can then use this correspondence to computationally encode free-format self-descriptions.

In Study 1, we assess the accuracy of this approach by investigating how well predicted values correspond to expert ratings. We begin by deriving an index of semantic similarity between a large number of words from word association data. We then predict words' values on the Big Five dimensions using their similarity towards words for which values on these dimensions are already known. Finally, the obtained scores are compared with expert ratings of the words' correspondence to the various domains.

In Study 2, we examine if we can use this technique to computationally encode the responses on a free-format personality test. First, participants complete a free-format and a fixed-format self-description personality assessment. For each participant, we create several personality profiles, each containing a score on openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. A first profile is generated by computationally encoding the responses of a free-format test using the technique described in Study 1. A second profile is obtained using expert's ratings of those responses. A final profile is acquired from a fixed-format personality inventory. The different personality profiles are then compared, to assess the performance of this approach of computationally coding free-format responses.

4.2 Study 1: predicting Big Five correspondence of words

In this first study, we investigate whether we can computationally estimate words' correspondence to the Big Five using their semantic similarity towards words for which this correspondence is already known. To that end, we make use of two freely available datasets: a large association dataset, from which we will derive semantic similarity; and a norming dataset, which contains correspondence to the Big Five for a large number of words.

4.2.1 Method

4.2.1.1 Materials.

Associative strength for a large number of words was obtained from the Dutch Small World of Words project, [7] a dataset comprising 5.56 million word associations collected in response to 16,000 cue words. To construct this dataset, each cue was presented to roughly 100 participants, who gave up to three responses per cue in what is called a continued word association task (see De Deyne, Navarro, & Storms, 2013, for full details[8]).

Expert ratings of the correspondence of 3,429 Dutch words and expressions to the Big Five dimensions were available through S.P.O.L.A.P. (1987), a norming dataset created to allow future administrators of the free-format self-description test to score obtained

---

[7] See `www.smallworldofwords.com`. Note that this project also contains word association datasets in several other languages.

[8] We made use of a more recent version of this dataset, which was somewhat larger than the published version but otherwise similar in all aspects.

responses without having to gather expert ratings. Items were obtained from the most frequent responses of 3,000 participants to the free-response self-description personality assessment. Each word or expression was rated by ten trained psychologists, who were asked to consider it a description of a person, and to subsequently judge to what extent that person would possess certain personality traits on a four-point scale: the person clearly possesses the trait, the person possesses the trait to some extent, the person does not really possess the trait, or the person does not possess the trait at all. Judged traits included conscientiousness, extraversion, agreeableness, and neuroticism. The fifth factor of the Big Five, generally interpreted as *openness to experience* in contemporary personality research, was instead defined as *general culture/education* in this work, which carries a rather different meaning. However, the researchers did collect ratings for the trait *creativity*; as this more closely resembles the construct openness to experience, we used these ratings for the final factor.

4.2.1.2 Procedure.

Semantic distance between each pair of the 16,000 cue words in the Small World of Words dataset was obtained through a cosine measure of similarity (Landauer & Dumais, 1997). In the context of word association data, this measure reflects the extent to which two words share associates: two words that elicit the exact same associations obtain a similarity of 1, while two words that share no associations obtain a similarity of 0. Previous research has indicated that this approach yields similarity estimates that show a high correspondence to human judgments of relatedness (De Deyne et al., 2013; De Deyne, Verheyen, & Storms, 2015).

The resulting similarity ratings were then used to inform two extrapolation methods. Both methods predict a word's correspondence to the Big Five dimensions using that word's similarity towards specific words for which values on those dimensions are already known.

The first extrapolation method we applied, *Orientation towards Paradigm Words*, predicts a word's score on some dimension using that word's similarity to certain paradigm words reflecting extreme values on that dimension (Kamps, Marx, Mokken, & de Rijke, 2004; Turney & Littman, 2003; Van Rensbergen et al., 2015). For each dimension, we selected four paradigm words based on instructions and trait descriptions in the S.P.O.L.A.P. norms (1987) and the Dutch NEO-PI-3 manual (Hoekstra, & De Fruyt, 2013), with two words corresponding to low values on that dimension, and two words corresponding to high values on that dimension (Table 4.1).

In the Orientation towards Paradigm Words method, a word's score on some dimension is first calculated as the sum of its (cosine) similarities towards both positive paradigm words, minus the sum of its similarities towards both negative paradigm words. This estimate is further refined by including the target word's similarity towards the nearest neighbors of each paradigm word: similarity towards neighbors of positive paradigm words is added to the estimate, while similarity to neighbors of negative paradigm words is subtracted from the estimate. In this context, nearest neighbors refer to, out of the 16,000 cue words, the words with the highest similarity towards that paradigm word. The number of neighbors of each paradigm word contributing to the estimate was included as parameter k, with k ranging from 0 to 500 similar to Van Rensbergen et al. (2015).

Table 4.1

English translation of the four paradigm words that were used for each of the Big Five dimensions (Dutch source)

Dimension	Positive Paradigm Words (*Dutch source*)	Negative Paradigm Words (*Dutch source*)
Openness to experience	creative, inventive (*creatief, inventief*)	objective, practical (*objectief, praktisch*)
Conscientiousness	orderly, punctual (*ordelijk*, *stipt*)	disorder, nonchalant (*wanorde*, *nonchalant*)
Extraversion	extrovert, smooth/easy (*extravert*, *vlot*)	introvert, silent (*introvert*, *zwijgzaam*)
Agreeableness	friendly, tolerant (*vriendelijk, tolerant*)	unfriendly, bossy (*onvriendelijk, bazig*)
Neuroticism	neurotic, nervous (*neurotisch, zenuwachtig*)	stable, relaxed (*stabiel, ontspannen*),

The second extrapolation method we employed, k-*Nearest Neighbors*, predicts ratings with the help of a limited set of expert ratings rather than rely on paradigm words (Bestgen & Vincze, 2012; Mandera, Keuleers, & Brysbaert, 2015; Recchia & Louwerse, 2015; Van Rensbergen et al., 2015). Here, the score of a word on some dimension is estimated as the mean score of its *k* nearest neighbors (assessed using cosine similarity) for which the rating on that dimension is known, which in our case refers to the mean rating of its *k* closest neighbors that are included in the S.P.O.L.A.P. (1987) norming dataset. These norms contain 857 words that are also part of the 16,000 cue words and as such can be used to contribute to the estimates, if they are among the *k* nearest neighbors of any of the words for which we extrapolate data. Values of the parameter for the neighborhood size *k* ranged from 1 to 500.

It may be of note to emphasize that under the *k*-Nearest Neighbors method, *k* refers to the number of neighbors of a target word, while under the Orientation towards Paradigm Words approach, *k* refers to the number of neighbors of each paradigm word. This difference also explains why *k*-Nearest Neighbors requires *k* to be at least 1, whereas Orientation towards Paradigm Words can function at $k = 0$, as this leaves similarity to the paradigm words themselves as a source of information.

4.2.2 Results

We predicted the openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism of the 16,000 cue words in the Small World of Words database, using both the Orientation towards Paradigm Words approach and the *k*-Nearest Neighbors technique. [9] Of these 16,000 words, 857 are contained in the S.P.O.L.A.P. (1987) norms and can be used to assess the accuracy of the two extrapolation methods. The remaining 2,584 items in the norm dataset were not part of the cue words, likely because they constituted mostly short sentences, a type of stimulus rarely used as cue in the association project.

The Orientation towards Paradigm Words method was run using values of k ranging between 0 and 500. Accuracy reached an optimum at $k = 25$, where extrapolated values showed a mean correlation of .63 with expert ratings (Table 4.2).

The *k*-Nearest Neighbors technique, in turn, was tested for values of k between 1 and 500. Accuracy peaked at $k = 10$, where trait predictions displayed a mean correlation of .78 with expert ratings (Table 4.2). Predictive accuracy of this method is systematically higher than that of Orientation towards Paradigm Words for all five factors; the difference is especially large for the factor openness to experience.

Overall, we find that both extrapolation methods yield estimates that show a high correspondence to expert ratings. As such, it seems feasible to use these techniques to computationally encode the responses of a free-format personality test.

---

[9] All ratings are made available at `https://osf.io/pmbvc/`

Table 4.2

Pearson correlations between expert ratings and estimates obtained with the k-*Nearest Neighbors and Orientation towards Paradigm Words extrapolation methods*

Extrapolation Method	O	C	E	A	N	Mean
Orientation towards Paradigm Words	.35	.74	.61	.85	.60	.63
k-Nearest Neighbors	.80	.81	.67	.87	.74	.78

Note. *n* = 857. O = Openness to experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism. Neighborhood parameter *k* was set to 10 for *k*-Nearest Neighbors, and to 25 for Orientation towards Paradigm Words. (Other values of *k* were tested as well, but yielded a lower mean accuracy.)

4.3 Study 2: computationally coding free-format responses

In this second study, we investigated how people's personality profiles generated by computationally coding responses correspond to profiles obtained by having experts rate the responses manually, and to profiles obtained through a standard, fixed-format personality questionnaire.

4.3.1 Method

4.3.1.1 Participants.

Seventy-one first-year psychology students (62 female) from the University of Leuven participated in return for course credit. Participants were native Dutch speakers, with ages ranging from 17 to 32 (M = 18.5).

4.3.1.2 Measures.

As a free-format personality test, we employed the free-response self-description method (Claeys et al., 1985), in which participants are asked to describe their own personality as completely as possible using any ten adjectives. It was stressed that they should describe how they really are, rather than how they want to be.

The Dutch version of the NEO-PI-3 (Hoekstra & De Fruyt, 2013) was used as a fixed-format personality test. Like the original, English NEO-PI-3 (McCrae et al., 2005), this test measures the Big Five dimensions openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism, through 240 items that participants respond to on a five-point Likert scale.

4.3.1.3 Materials.

As in Study 1, associative strength for a large number of Dutch words was available through the Dutch Small World of Words project (De Deyne et al., 2013).

Expert ratings of the score of 3,429 Dutch words on the Big Five domains were obtained from S.P.O.L.A.P. (1987); see the Materials section of Study 1 for more details.

4.3.1.4 Procedure.

Participants completed both the free-response self-description test and the NEO-PI-3 in a single session. The free-response assessment always preceded the NEO-PI-3.

Four personality profiles where created for each participant. A first profile was obtained by coding the responses of the free-response test using the S.P.O.L.A.P. expert ratings. A participant's score on a dimension was calculated as the mean of the norm ratings of their responses on that dimension. When a response was not included in the norms, but a synonym of that adjective was, the ratings of the synonym were used instead. When an adjective was not listed in the norms and no suitable synonym was found, the response was omitted. This approach has been found to be a valid method to obtain trait scores (S.P.O.L.A.P., 1987; Van den Broucke et al., 1989). 89% of responses were contained in the norms directly; after replacing 44 responses with synonyms, ratings were available for a total of 95% of responses. As an example of synonym replacement, the score of *shy (verlegen)* was used for the response *timid* (*timide)*, as the latter was not included in the norm words.

Two computationally encoded personality profiles were constructed using the approach described in Study 1. In short, a cosine measure of similarity was calculated between all word-pairs of the 16,000 cue words in the Small World of Words association dataset. Using these similarity indices, two extrapolation methods predicted the Big Five ratings of participants' responses to the free-format self-description task.

The first method, Orientation towards Paradigm Words, predicted scores using words' similarity towards the paradigm words listed in Table 4.1. The neighborhood parameter *k*, which indicates the number of nearest neighbors of each paradigm word that contribute to the estimates, was set at 25, the value with the highest accuracy in Study 1.

The second method, *k*-Nearest Neighbors, estimated the rating of a word as the mean rating of its *k* nearest neighbors (assessed through cosine similarity) that are included in the S.P.O.L.A.P. norms. Neighborhood parameter *k* was set to 10, the optimal value reported in Study 1.

The extrapolation methods could only predict ratings of responses contained in the 16,000 cue words of the association dataset. Initially, this constituted 93% of the obtained responses. Similarly to the approach used for coding the responses using expert ratings, we replaced 32 responses that were not contained in the list of cue words with a synonym that was; as a result, we were able to predict the rating of 98% of the responses (i.e., 3% more than in the direct S.P.O.L.A.P. coding). The remaining 2% of adjectives, words not part of the cue list for which no suitable synonym was found, were omitted.

A final personality profile was created from the NEO-PI-3. Scale scores for each dimensions were calculated based on the responses of the corresponding items.

4.3.2 Results

Table 4.3 lists the Pearson correlations between trait scores of the four personality profiles. Results show that encoding free-format responses computationally yielded scores very similar to having those responses coded by experts, with a mean correlation of .71 for the Orientation towards Paradigm Words extrapolation method, and .84 for the *k*-Nearest Neighbors technique. *k*-Nearest Neighbors outperformed Orientation towards Paradigm Words on all dimensions except agreeableness; similarly to the findings in Study 1, the difference is especially notable for the dimension openness to experience.

We also investigated to what extent participants' scores on the free-format self-description test correspond to their scores on the NEO-PI-3. We found a mean trait correlation of .40 when free-format responses are encoded by experts, .37 when they are encoded using Orientation towards Paradigm Words, and .41 when they are encoded using *k*-Nearest Neighbors (Table 4.3). Examining correspondence across the five tested dimension, we found the highest correlations for extraversion, followed by conscientiousness, agreeableness, and neuroticism. Correlations were very low for openness to experience, and even negative when responses are encoded computationally.

Table 4.3

Pearson correlations between Big Five domain scores of participants obtained from encoding the free-format self-description test using three different methods, and from the NEO-PI-3.

Comparison	O	C	E	A	N	Mean
Expert Ratings vs Paradigm Words	.33	.86	.76	.89	.71	.71
Expert Ratings vs *k*-Nearest Neighbors	.82	.90	.77	.87	.82	.84
NEO-PI-3 vs Expert Ratings	.08	.46	.56	.43	.47	.40
NEO-PI-3 vs Paradigm Words	-.11	.46	.59	.45	.47	.37
NEO-PI-3 vs *k*-Nearest Neighbors	-.04	.48	.60	.44	.59	.41

Note. n = 71. O = Openness to experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism. Expert Ratings refers to the free-format self-description test encoded by experts. Paradigm Words refers to that test coded using the Orientation towards Paradigm Words method, at k = 25. k-Nearest Neighbors refers to that test coded using the k-Nearest Neighbors method, at k = 10.

4.4 General Discussion

A problematic issue when deriving personality from free-format self-descriptions is the matter of encoding obtained responses. Generally, they are rated by experts, a time-consuming and expensive procedure. In the current paper, we proposed two methods to computationally code these responses. Both approaches derive similarity indices from word association data, and estimate the correspondence of responses to the Big Five domains using those responses' similarity towards words for which this correspondence is already known.

In Study 1, we find that both methods produce estimates that show high correlations to expert ratings. In Study 2, we had participants complete a free-response self-description personality test, in which they described their personality using any ten words. Correspondence of obtained responses to the Big Five was encoded both by experts and using the two extrapolation methods, revealing very similar trait scores.

When comparing the two extrapolation methods we describe, we find a slightly higher accuracy for the *k*-Nearest Neighbors approach. In a previous study in which we predicted the valence, arousal, and dominance of words, we also reported a higher performance for this method (Van Rensbergen et al., 2015). This is line with expectations, as the estimates of this technique are based directly on the human ratings with which accuracy is assessed (note, though, that expert ratings of some word do not contribute to that word's predicted value). At the same time, this reliance on expert ratings highlights the downside of the *k*-Nearest Neighbors method: it can only function when human judgments are already available for a number of words. Generally, a relatively small sample of words is sufficient to make accurate predictions; for example, in the current research Big Five correspondence was successfully predicted using expert ratings of 'only' 857 words, while in a previous study, we were able to successfully estimate valence, arousal, and dominance scores using human ratings of 1,000 words. Nevertheless, the need for an existing set of expert ratings is clearly a drawback of the *k*-Nearest Neighbors approach, as it enforces a strict limit on which measures it can predict. The Orientation towards Paradigm Words technique does not suffer from this issue; here, the only requirement is choice of paradigm words.

When we examine how well estimates on the various domains correspond to expert ratings, we find that predictive accuracy is by far the lowest when predicting openness to experience with the Orientation towards Paradigm Words method. This could be a result of an inadequate choice of paradigm words; alternatively, it could simply be an effect of the elusive nature of this factor, which has received various other interpretations by different researchers, including culture (Tupes & Christal, 1961), intellect (Peabody & Goldberg, 1989) or intelligence (Borgatta, 1964), or refinement (Smith, 1967).

Taken together, we find that when one is interested in deriving personality from free-format tests, computationally encoding the responses can be a viable approach. So far, responses were coded either directly by experts, or indirectly through norm datasets such as that of S.P.O.L.A.P. (1987). Compared with the former, our method takes considerably less time and effort. Compared with the latter, our technique gives access to ratings for a much larger set of words than any available dataset. When using the Orientation towards Paradigm Words approach, two additional advantages come to light. As the predictions of this technique do not rely on expert ratings, it provides a way around the issue of bias in expert's encoding of responses. Second, this method is not limited to predicting Big Five values of words, but grants easy access to estimates on any measure, as long as suitable paradigm words can be found among the cue words in the association dataset.

In the current study, we computationally encoded the responses of a free-format self-description personality test. The technique we followed could easily be applied to predict personality from word use in other contexts where people describe themselves, such as social media posts or diary entries. Future research should investigate how well this method performs compared with previous attempts at computationally deriving personality from this type of information (e.g., Park et al, 2014).

Under the approach we outlined, predicted values are based on similarity indices derived from word association data. An alternate approach to obtaining information about relatedness makes use of word co-occurrence in text corpora. Here, the likelihood of two words co-occurring in some piece of text is taken as a measure of the similarity between the two words. While this method has had considerable success in the literature (e.g., Bullinaria & Levy, 2007; Church & Hanks, 1990; Landauer & Dumais, 1997), it should be note that the resulting similarity indices are often less accurate than similarity estimates derived from word association data (De Deyne, Peirsman, & Storms, 2009; De Deyne et al., 2015; see also Van Rensbergen et al., 2015). Nevertheless, when working in languages where access to word associations datasets is limited but large text corpora are available, this alternative approach may be preferable.

A secondary goal of our research was to examine the correspondence between the free-response self-description personality test, and the more traditional, fixed-format NEO-PI-3 (McCrae et al., 2005) questionnaire. We find a moderate overlap between the personality profiles of both tests, with a mean correlation of .40 when free-format responses were coded by experts, and .37 and .41 when they were encoded computationally. These correlations are lower than expected, as the study that originally proposed the free-response self-description measure reported a mean correlation of .59 with a fixed-format personality inventory (Claeys et al., 1985).

This discrepancy likely involves the factor openness to experience, as the NEO-PI-3 and the free-format test show almost no correspondence on this domain. An explanation may lie with this domain being interpreted as *creativity* rather than *openness to experience* in the S.P.O.L.A.P. research, as all free-format scores are based on these norms either directly (when response coding is based on expert ratings or *k*-Nearest Neighbors predictions) or indirectly (when using Orientation towards Paradigm Words, as choice of paradigm words was based in part on the instructions used to gather the S.P.O.L.A.P. norms). Future research could investigate this issue by gathering expert ratings for the correspondence of words to *openness to experience*, and examine whether using these new ratings to code the free-format test improves correspondence with the NEO-PI-3.

Even when disregarding the factor openness to experience, the free-format – fixed-format correspondence we report is somewhat lower than what was described by Claeys et al. (1985). This could be a result of the used fixed-format measure: we made use of the contemporary NEO-PI-3, a 240-item test where participants respond on a five-point Likert scale, while Claeys and colleagues administered the Dutch Five Personality Factors Test (Elshout & Akkerman, 1975), a currently outdated 70-item questionnaire where participants respond on a seven-point Likert scale. Alternatively, this could be caused by the age of the expert ratings we used; as they were gathered almost 30 years ago, they are likely somewhat out-of-date. However, as we are able to successfully predict these ratings with the orientation towards paradigm words method (which does not make use of the ratings), this does not seem like a substantial problem.

In any case, as the free-format self-description and the NEO-PI-3 supposedly measure the same underlying construct (the Big Five), ideally they would display a stronger correspondence than what we report. Of course, this moderate correspondence does not necessarily indicate an issue with either test; for example, it is possible that both tests measure a different aspect of the same construct. Existing research reports a similar to higher validity for the free-format test than for fixed-format measures (Claeys et al., 1985; S.P.O.L.A.P., 1987; Van Den Broucke et al., 1989), although again, these studies where performed with questionnaires that are currently outdated. Future research should address this issue by using contemporary tools to compare the ability of both types of test to predict (social) behavior.

References

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, *9*(3), 272–279. doi:10.1037/h0025907

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, *21*(3), 372–374. doi:10.1177/0956797609360756

Bestgen, Y., & Vincze, N. (2012). Checking and bootstrapping lexical norms by means of word similarity indexes. *Behavior Research Methods*, *44*(4), 998–1006. doi:10.3758/s13428-012-0195-z

Borgatta, E. F. (1964). The structure of personality characteristics. *Behavioral Science*, *9*(1), 8–17. doi:10.1002/bs.3830090103

Broucke, S. Van Den, Soete, G. De, & Böhrer, A. (1989). Free-response self-description as a predictor of success and failure in adolescent exchange students. *International Journal of Intercultural Relations*, *13*, 73–91. doi:10.1016/0147-1767(89)90036-9

Bullinaria, J., & Levy, J. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526. doi:10.3758/BF03193020

Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29.

Claeys, W., De Boeck, P., Van Den Bosch, W., Biesmans, R., & Böhrer, A. (1985). A comparison of one free-format and two fixed-format self-report personality assessment methods. *Journal of Personality and Social Psychology*, *49*(4), 1028–1039. doi:10.1037/0022-3514.49.4.1028

Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory: Manual, form S and form R*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, *64*(1), 21–50. doi:10.1207/s15327752jpa6401_2

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–98. doi:10.3758/s13428-012-0260-7

De Deyne, S., Navarro, D. J., & Storms, G. (2015). Associative strength and semantic activation in the mental lexicon: evidence from continued word associations. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*.

De Deyne, S., Peirsman, Y., & Storms, G. (2009). Sources of semantic similarity. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (pp. 1834–1839). Austin, TX: Cognitive Science Society.

De Deyne, S., Verheyen, S., & Storms, G. (2015). The role of corpus size and syntax in deriving lexico-semantic representations for a wide range of concepts. *Quarterly Journal of Experimental Psychology*, *68*(8), 1643–1664. doi:10.1080/17470218.2014.994098

Deese, J. (1965). *The structure of associations in language and thought*. Baltimore, MD: Johns Hopkins University Press.

Elshout, J. J., & Akkerman, A. E. (1975). Vijf Persoonlijkheidsfactoren Test (5PFT) Handleiding [Five Personality Factors Test Manual]. Nijmegen, The Netherlands: Berkhout.

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from twitter. *IEEE International Conference on Social Computing*, 149–156. doi:10.1109/PASSAT/SocialCom.2011.33

Goldberg, L. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26. doi:10.1037/0003-066X.48.1.26

Hoekstra, H. A., & De Fruyt, F. (2013). *NEO-PI-3 Persoonlijkheidsvragenlijst [NEO-PI-3 personality inventory].* Manuscript in Preparation.

Iacobelli, F., Gill, A. J., Nowson, S., & Oberlander, J. (2011). Large scale personality classification of bloggers. In S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective Computing and Intelligent Interaction* (pp. 568–577). Berlin, Germany: Springer. doi:10.1007/978-3-642-24571-8_71

Kamps, J., Mokken, R. J., Marx, M., & De Rijke, M. (2004). Using WordNet to measure semantic orientation of adjectives. In *Proceedings of LREC* (Vol. 4, pp. 1115–1118). doi:10.1.1.134.483

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. doi:10.1037/0033-295X.104.2.211

Mandera, P., Keuleers, E., & Brysbaert, M. (2015). How useful are corpus-based methods for extrapolating psycholinguistic variables? *The Quarterly Journal of Experimental Psychology*, *68*(8), 1623–1642. doi:10.1080/17470218.2014.988735

McCrae, R., Costa, P., & Martin, T. (2005). The NEO–PI–3: A more readable revised NEO personality inventory. *Journal of Personality Assessment*, *84*(3), 37–41. doi:10.1207/s15327752jpa8403

McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and associative relations in adolescents and young adults: Examining a tenuous dichotomy. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), *The adolescent brain: Learning, reasoning, and decision making* (pp. 39–66). Washington, DC: American Psychological Association. doi:10.1037/13493-002

Mollin, S. (2009). Combining corpus linguistic and psychological data on word co-occurrences: Corpus collocates versus word associations. *Corpus Linguistics and Linguistic Theory*, *5*(2), 175–200. doi:10.1515/CLLT.2009.008

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–7. doi:10.3758/BF03195588

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: replicated factors structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, *66*(6), 574–583. doi:10.1037/h0040291

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... Seligman, M. E. P. (2015). Automatic Personality Assessment Through Social Media Language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. doi:10.1037/pspp0000020

Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology*, *57*(3), 552–567. doi:10.1037/0022-3514.57.3.552

Peng, K., Nisbett, R. E., & Wong, N. Y. C. (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods*, *2*(4), 329–344. doi:10.1037/1082-989X.2.4.329

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296–1312. doi:10.1037/0022-3514.77.6.1296

Pervin, L. (1976). A free-response description approach to the analysis of person-situation interaction. *Journal of Personality and Social Psychology*, *34*(3), 465–474. doi:10.1037/0022-3514.34.3.465

Pervin, L. A., & John, O. P. (1999). *Handbook of personality: Theory and research*. New York, NY: The Guilford press.

Poortinga, Y. H. (1989). Equivalence of cross-cultural data: an overview of basic issues. *International Journal of Psychology*, *24*(6), 737–756. doi:10.1080/00207598908247842

Potkay, C. R., & Allen, B. P. (1973). The adjective generation technique: an alternative to adjective check lists. *Psychological Reports*, *32*(2), 457–458. doi:10.2466/pr0.1973.32.2.457

Recchia, G., & Louwerse, M. M. (2015). Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *Quarterly Journal of Experimental Psychology*, *68*(8), 1584–1598. doi:10.1080/17470218.2014.941296

S.P.O.L.A.P. (1987). *Vrije zelfbeschrijvingsmethode [Free-response self-description method]*. Lisse, the Netherlands: Swets & Zeitlinger.

Schönbach, P. (1972). Likableness ratings of 100 German personality-trait words corresponding to a subset of Anderson’s 555 trait words. *European Journal of Social Psychology*, *2*(3), 327–333. doi:10.1002/ejsp.2420020309

Szalay, L. B., & Deese, J. (1978). Subjective meaning and culture: An assessment through word associations. Hillsdale, NJ: Lawrence Erlbaum.

Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings (Tech. Rep. No. ASD-TR-61-97)*. Lackland Air Force Base, TX: U.S. Air Force.

Turney, P., & Littman, M. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, *21*(4), 315–346. doi:10.1145/944012.944013

Van Rensbergen, B., De Deyne, S., & Storms, G. (2015). Estimating affective word covariates using word association data. *Behavior Research Methods*. Advance online publication. doi: 10.3758/s13428-015-0680-2

Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373. doi:10.1016/j.jrp.2010.04.001

Chapter 5

Measuring brand personality

Abstract

Brand personality, the human traits or characteristics associated with a brand, is a commonly used approach to assess brand image or brand differentiation. Traditionally, brand personality is measured with direct ratings, where participants give a score of the correspondence of the brand to various personality traits. One major drawback of this method is that these ratings are very vulnerable to response bias and social desirability effects. In the current study, we propose deriving brand personality from the associations given to a brand in a word association task. This approach holds two large advantages over direct ratings. First, participants in an association task are offered very little chance at cognitive monitoring, making the resulting data very resistant against response bias. Second, the proposed approach offers a huge number of brand personality indices despite requiring only one session of data collecting. We used word associations to six brands to predict the brands' correspondence to the brand personality domains responsibility, activity, aggressiveness, simplicity, and emotionality. The resulting estimates are then compared to direct ratings, revealing a strong correlation for activity and aggressiveness, a moderate correlation for emotionality, and a low and insignificant correlation for responsibility and simplicity. We discuss some limits to the approach we followed and offer suggestions as to how future research could address these.

5.1 Introduction

Brand differentiation is an important indicator of firm performance (Madden, Fehle, & Fournier, 2006). One approach of assessing the distinctiveness of a brand is through brand personality, the set of human traits or characteristics associated with a brand (Aaker, 1997).

Similarly to human personality, brand personality is generally measured on a set of standardized dimensions. The most commonly used approach entails the dimensions sincerity, excitement, competence, sophistication, and ruggedness, each with a number of sub-facets; a factor analysis showed that these aspects were the most powerful in differentiating the perception of different brands (Aaker, 1997). Although widely used, recent research has uncovered some theoretical issues with this scale, including unclear construct validity (e.g., some of the facets do not refer to personality traits but rather to respondent characteristics such as age or gender) and cross-cultural non-replicability (Azoulay & Kapferer, 2003; Bosnjak, Bochmann, & Hufschmidt, 2007). In response, an extensive study was undertaken to develop a scale of brand personality that consisted of only personality items and showed cross-cultural validity (Geuens, Weijters, De Wulf, 2009). This led to the dimensions responsibility, activity, aggressiveness, simplicity, and emotionality, again all with a number of sub-facets.

Traditionally, brand personality is measured by having participants rate brands on their correspondence to the various traits. While this is a simple and straightforward procedure, direct ratings have the drawback of being very vulnerable to cognitive monitoring on behalf of the participant, as evidenced by response bias or social desirability effects (Edwards, 1957; Furnham, 1986; Nederhof, 1985). While all methods of measuring subjective aspects share this issue to some extent, most have mechanisms to defend against it, such as obfuscating what is being measured. In contrast, measuring brand personality through direct ratings offers no effective way of combatting response bias. In other contexts, word associations have proven a useful tool to investigate subjective meaning (Deese, 1965). Word association data is collected by having participants write down the first words that come to mind after reading a cue word. The probability that a certain word is given in response to that cue can then be considered a measure of how strongly associated people consider the two words to be (De Deyne, Navarro, & Storms, 2013; Nelson, McEvoy, & Schreiber, 2004).

When measuring subjective aspects such as attitude, word associations offer several advantages over direct ratings. First, subjective information is often below the level of awareness, and cannot always be verbalized in response to direct questions; word associations frequently allow access to these subconscious attitudes (Szalay & Deese, 1978). Additionally, participants in an association task are offered little chance for conscious monitoring, and the measure itself can be considered semi-implicit; giving three words that come to mind is presumably less overt than explicitly asking for an affective rating. Implicit measures are especially useful when examining attitudes on controversial topics (Greenwald & Banaji, 1995). While there are several other methods to measure implicit attitudes, word associations have the advantage of being straightforward to collect. This combination of minimal cognitive monitoring and being semi-implicit make the associative method quite resistant against response bias and social desirability (Szalay & Deese, 1978).

Recently, we have used word association data to predict the valence, arousal, and dominance of words (Van Rensbergen, De Deyne, &

Storms, 2015), as well as words' correspondence to the Big Five dimensions: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (Van Rensbergen, Kuppens, Storms, & De Deyne, 2015). Obtained estimates were compared to human ratings, revealing correlations of up to .91, .84, and .85 for valence, arousal, and dominance, and up to .80, .81, .67, .87, and .74 for the Big Five.

In this paper, we propose following a similar approach to measure brand personality. We start by having participants give associations towards a number of brands, and add these to an existing word association dataset. From these associations we then derive the semantic similarity between the brands and various brand personality traits; this similarity is taken as an index of the brand's score on those traits. Finally, we compare obtained trait scores with direct ratings.

5.2 Method

5.2.1 Materials

To assist with calculating the similarity between brands and brand traits, we make use of the Dutch Small World of Words word association dataset, which contains 5.56 million word associations collected in response to about 16,000 cue words. Each cue was present to roughly 100 participants, who responded with up to three associations per cue (see De Deyne et al., 2013, for full details[10]). This dataset includes all trait names of the brand personality scale of Geuens et al. (2009); only a direct translation of the trait *bold* is not included (see Table 5.1).

---

[10] We made use of an updated version of this dataset, larger than the published version but otherwise similar in all aspects.

Table 5.1

The brand personality scale of Geuens et al. (2009) (Dutch translation we used)

Dimension	Traits (*Dutch translation*)
Responsibility	down to earth, stable, responsible (*nuchter, stabiel, verantwoordelijk*)
Activity	active, dynamic, innovative (*actief, dynamisch, innovatief*)
Aggressiveness	aggressive, bold (*agressief, brutaal*[a])
Simplicity	ordinary, simple (*alledaags, eenvoudig*)
Emotionality	romantic, sentimental (*romantisch, sentimenteel*)

[a]*brutaal* is not a direct translation of *bold*; however, a closer match was not found among the cue words in the word association dataset.

5.2.2 Participants

Participants were 283 first year psychology students (232 female) from the University of Leuven, with ages ranging from 17 to 56 (M = 19). All participants were native Dutch speakers and took part in return for course credit.

5.2.3 Brands

We examined six brands: Armani, Coca Cola, Colgate, Google, Harley Davidson, and McDonald's. These were selected as they would be well-known by all participants, and intuitively show relatively large differences in terms of brand image.

5.2.4 Procedure

5.2.4.1 Data collection.

Each participant was assigned three brands, selected at random from the six brands we investigated.

Participants first completed a word association task. They were presented with a cue word and were asked to respond with the first three words that came to mind. Each participant was presented with a total of 15 cue words: three brand names and 12 filler items (selected at random per participant, from a list of 30 common words). Brand cues were always presented as the fourth, eighth, and twelfth item. Participants had the option of skipping to the next cue if they did not know the current cue word or had no further associations.

After this, the same participants completed a brand personality questionnaire. For each of the three brands they were presented in the word association task, they were asked to consider to what extent the 12 traits (see Table 5.1) of the brand personality scale of Geuens et al. (2009) applied. Participants responded on a scale ranging from 1 (*not characteristic for the brand at all*) to 7 (*very characteristic for the brand*).

5.2.4.2 Measuring brand personality.

In order to be able to estimate the semantic similarity between the brands and the various brand personality traits, we first added the obtained brand associations to Small World of Words dataset. We then calculated the cosine similarity (Landauer & Dumais, 1997) between each combination of the roughly 16,000 cue words of this combined dataset (i.e., including the brand names). In the context of word association data, a cosine measure reflects the extent to which two words overlap in associations: words that always elicit the exact same associates obtain a value of 1, while words that have no associates in common obtain a value of 0. Research has indicated that this measure shows a high correlation to human judgments of relatedness (De Deyne et al., 2013; De Deyne, Verheyen, & Storms, 2015).

The resulting similarity indices were then used to measure the correspondence between brand names and personality traits. A first index of a brand's correspondence to a trait was simply the cosine similarity between the two. This measure was then refined by including the brand's similarity towards the nearest neighbors of the trait, that is, out of the 16,000 words, the words with the highest cosine similarity to that trait. The number of neighbors of the traits that contributed to the measure was included as parameter k, where k ranged from 0 to 500, similar to previous research where affective word covariates were estimated (Van Rensbergen, De Deyne, et al., 2015; Van Rensbergen, Kuppens, et al., 2015). The brand's final score on the trait was computed as the sum of the cosine similarities of the brand towards the trait and towards the trait's k nearest neighbors. A brand's score on each of the five dimensions of the brand personality scale was then simply the mean of its scores on the traits corresponding to that dimension.

Finally, to assess how these measures compare with direct human ratings, we obtained a second set of brand personality scores from direct ratings; a brand's score on each of the five dimensions was calculated as the mean of participants' ratings on all traits corresponding to that dimension.

5.3 Results

Figure 5.1 displays the correspondence between direct brand personality ratings and brand personality assessed through the cosine similarity between a brand's name and the various traits and their nearest neighbors, with neighborhood size parameter k set to 2 (the value for which overall correspondence is the highest, see also Table 5.2). Each data point reflects a brand's score on one trait, resulting in three points per brand for the dimensions responsibility and activity, and two points per brand for the dimensions aggressiveness, simplicity, and emotionality. For all five dimensions, regression lines suggest a positive correlation between the two measures of brand personality. Correspondence is particularly strong for the dimension of aggressiveness, where all data points fall very close to the secondary diagonal.

Pearson correlations between scores obtained using the two methods (Table 5.2) confirm the very high correspondence on aggressiveness, with correlations of up to .997. Correspondence is also very high for activity, with correlations of up to .92, and for emotionality, with correlations of up to .82. Correspondence is lower for the dimensions of responsibility and simplicity, where correlations peak at .52 and .42, respectively.

We find that for all dimensions except aggressiveness, the value of k (the number of nearest neighbors of each trait contributing to the semantic similarity score) has a strong effect on the correspondence between the two methods. Moreover, for these four dimensions, correspondence shows a multimodal distribution (i.e., multiple peaks) with respect to parameter k, in contrast to existing research where affective word covariates were estimated using a similar method (Recchia & Louwerse, 2015; Van Rensbergen, De Deyne, et al., 2015; Van Rensbergen, Kuppens, et al., 2015). As there is no a priori reason to expect this parameter to behave differently when measuring brand personality, perhaps more data points are needed to reveal the true nature of how correspondence varies with respect to k.

When investigating the statistical significance of the correspondence between the two methods, we had to take into account that for each dimension, we only had access to six data points, one per brand. [11] It has been suggested that when sample sizes are very small, significance should be assessed through randomization tests (also called permutation tests or exact tests), rather than t-tests (Edgington & Onghena, 2007; Ludbrook & Dudley, 1998). In this context, a randomization test first calculates what correlations would be expected if there was no correspondence between the two tests by calculating all possible correlations obtained by permuting the six data points (i.e., it

[11] Note that an analysis at the level of respondents is not possible: a reliable cosine similarity requires associations of multiple persons, so this method only yields aggregate scores.

finds the distribution of the test statistic under the null hypothesis). Statistical significance is then obtained from the proportion of these correlations that is lower than the empirically found correlation. This process was applied for each measure, for each tested value of k; the resulting levels of significance are marked in Table 5.2 with asterisks. We find that the two methods show a significant correspondence on aggressiveness for all values of k, and on activity for most k's. Emotionality only shows significant correlations at low values of k, and the correlations of responsibility and simplicity never achieve significance.

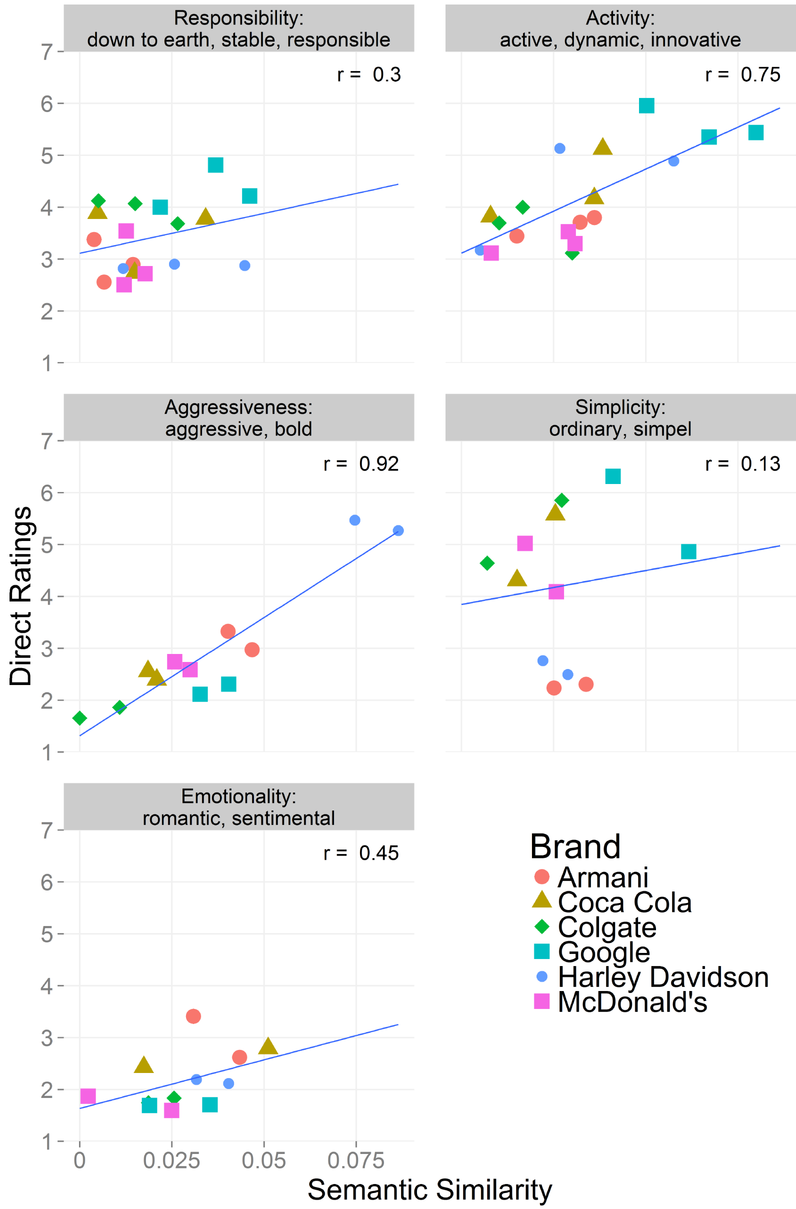

Figure 5.1. Scatterplots (one data point per brand per trait) and regression lines of correspondence between brand personality measured through direct ratings and through semantic similarity assessed with word association data, with neighborhood size parameter *k* set to 2 (the value with highest mean correspondence to direct ratings, see table 5.2).

Table 5.2

Pearson correlations between direct brand personality ratings and brand personality measured through semantic similarity

k	Resp.	Activity	Aggr.	Simplicity	Emot.
0	.33	0.72	.95*	-.21	.82*
1	.39	0.79	.94*	.13	.76*
2	.52	.91*	.95**	.16	.77*
5	.51	0.85	.99**	.32	.55
10	.51	.89*	.997**	.41	.37
15	.48	.89*	.99**	.32	.16
20	.49	.92*	.99**	.29	.08
25	.49	.92*	.99**	.28	.01
50	.51	.92*	.99**	.29	-.04
100	.44	.92*	.99**	.32	-.04
250	.51	.92*	.99**	.17	.21
500	.49	.91*	.97**	.09	.32

Note. n = 6. Resp. = Responsibility, Aggr. = Aggressiveness, Emot. = Emotionality. k refers to the neighborhood size parameter used for calculating semantic similarity. Statistical significance was assessed with a randomization test.

$^{*}p < .05$, $^{**}p < .01$

5.4 Discussion

We have outlined a novel approach of measuring brand personality, by using the overlap in word associations between brands and brand personality traits as an index of the brands value on those traits.

Traditionally, brand personality is obtained through direct ratings. In comparison, the method we propose holds several advantages. First, it can be considered a semi-implicit measure, as the word association data on which our method is based shows remarkable resistance against response bias and social desirability effects (Szalay & Deese, 1978). This property makes association data well suited for any measure of attitude or opinion, especially so in contexts where the topic being measured could be considered somewhat controversial (e.g., perception of banks in times of financial crisis).

The method we outline also has a huge practical advantage in that it allows choosing the personality traits that will be measured *after* data collection: once associations towards the brand have been collected, the brand's semantic similarity towards any of the thousands of cue words in the association database can be calculated. Of course many of these words are not related at all to brand personality, yet they likely contain most traits of interest to brand personality researchers. For example, most of the traits of the Aaker (1997) brand personality scale are included in the word association dataset we used; as such, we could easily calculate the scores on this scale for the brands we investigated, without collecting additional data.

Measuring brand personality with word associations also has a number of disadvantages. Perhaps the most obvious drawback is that it requires access to word association data; creating a word association dataset is definitely a large undertaking. Fortunately, such databases already exist in many languages. For example, next to the Dutch dataset we used, the Small World of Words project contains datasets in English, French, German, Spanish, Rioplatense Spanish, Vietnamese, Japanese, and Cantonese.

A separate problematic issue with the approach we have outlined concerns the unclear validity of the resulting brand personality scores. While we found that they generally correlated positively with direct ratings, this correspondence was not always significant. Of course, this does not necessarily indicate that assessing brand personality through associations is not a valid measure; as indicated previously, direct ratings are far from a perfect method of measuring opinion as they are very sensitive to response bias. Moreover, the explanatory power of the correlations we reported is rather limited, due to the very small sample size on which they are based. This can be seen, for example, in how they displayed a multimodal distribution with respect to neighborhood size parameter k; in previous research where other subjective word covariates were predicted with much larger sample sizes, correspondence with human ratings always showed unimodal distributions.

To address these issues, future research should investigate a much larger range of brands. This would allow for a more reliable measure of the correspondence between predictions and direct ratings, and would help alleviate any sampling bias in choice of brands; the six brands we examined may not be very representative of all brands. Additionally, future research should definitely assess to what extent brand personality derived from word association data corresponds to more qualitative measures of brand image, such as in-depth interviews.

References

Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research*, *34*(3), 347–356.

Azoulay, A., & Kapferer, J.-N. (2003). Do brand personality scales really measure brand personality? *Journal of Brand Management*, *11*(2), 143–155. doi:10.1057/palgrave.bm.2540162

Bosnjak, M., Bochmann, V., & Hufschmidt, T. (2007). Dimensions of brand personality attributions: a person-centric approach in the German cultural context. *Social Behavior and Personality: An International Journal*, *35*(3), 303–316. doi:10.2224/sbp.2007.35.3.303

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–98. doi:10.3758/s13428-012-0260-7

De Deyne, S., Verheyen, S., & Storms, G. (2015). The role of corpus size and syntax in deriving lexico-semantic representations for a wide range of concepts. *Quarterly Journal of Experimental Psychology*, *68*(8), 1643–1664. doi:10.1080/17470218.2014.994098

Deese, J. (1965). *The structure of associations in language and thought*. Baltimore, MD: Johns Hopkins University Press.

Edgington, E., & Onghena, P. (2007). *Randomization tests*. Boca Raton, FL: CRC Press.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Ft Worth, TX: Dryden Press.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, *7*(3), 385–400. doi:10.1016/0191-8869(86)90014-0

Geuens, M., Weijters, B., & De Wulf, K. (2009). A new measure of brand personality. *International Journal of Research in Marketing*, *26*(2), 97–107. doi:10.1016/j.ijresmar.2008.12.002

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. doi:10.1037/0033-295X.104.2.211

Ludbrook, J., & Dudley, H. (1998). Why permutation tests are superior to t and f tests in biomedical research. *The American Statistician*, *52*(2), 127. doi:10.2307/2685470

Madden, T. J., Fehle, F., & Fournier, S. (2006). Brands matter: an empirical demonstration of the creation of shareholder value through branding. *Journal of the Academy of Marketing Science*, *34*(2), 224–235. doi:10.1177/0092070305283356

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*(3), 263–280. doi:10.1002/ejsp.2420150303

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–7. doi:10.3758/BF03195588

Recchia, G., & Louwerse, M. M. (2015). Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *Quarterly Journal of Experimental Psychology*, *68*(8), 1584–1598. doi:10.1080/17470218.2014.941296

Szalay, L. B., & Deese, J. (1978). *Subjective meaning and culture: An assessment through word associations*. Hillsdale, NJ: Lawrence Erlbaum.

Van Rensbergen, B., De Deyne, S., & Storms, G. (2015). Estimating affective word covariates using word association data. *Behavior Research Methods*. Advance online publication. doi: 10.3758/s13428-015-0680-2

Van Rensbergen, B., Kuppens, P., Storms, G., & De Deyne, S. (2015). *Computationally coding responses of a free-format self-description personality test using word association data*. Manuscript submitted for publication.

Chapter 6

General conclusion

In this dissertation, we set out to investigate the relation between a word's affective meaning and its connectivity in the mental lexicon. In four empirical studies, we have presented clear evidence that words that are connected in the mental lexicon share a large number of affective attributes, and that this correspondence is strong enough to predict a word's affective values from its connectivity to words for which these values are already known. Taken together, this shows that affective aspects are undeniably involved with the structure of the mental lexicon.

In this final chapter, we'll first give a short summary of each empirical study. We'll then address some of the limitations of the presented research and offer several suggestions for future research into the role of affectivity in the mental lexicon.

6.1 Summary of empirical studies

In Chapter 2, we discovered that linked words share affective and perceptual attributes. Combining word association data with several norming datasets, we found that when presented with a cue word, participants respond with associations that are alike in terms of valence, arousal, dominance, as well as in terms of concreteness, a non-affective factor that has previously been found to be related to affective meaning (e.g., Vigliocco, Meteyard, Andrews, & Kousta, 2009). Existing research reported similar findings, both when using word association data (e.g., Cramer, 1968; Pollio, 1964; Staats & Staats, 1959) and in other fields (e.g., Klauer, 1997; Bleasdale, 1987). Compared with these studies, the approach we followed provided multiple advantages. First, in contrast to this existing literature, we investigated a very large number of words, making generalization towards the entire mental lexicon feasible. Additionally, by examining the multiple dimensions simultaneously, we were able to ascertain that each dimension showed an independent effect; that is, each investigated factor showed a significant cue-response correspondence, independently of cue and response values on the remaining factors. Finally, our approach allowed us to assess the relative importance of each dimension, finding that cue-response correspondence is by far the highest for the factor valence.

In sum, this chapter showed that words people deem to be related tend to share affective attributes. Considering this, it seems reasonable to assume that one can obtain information about a word's affective meaning from the words to which it is connected. We examined this assumption in the next three chapters, by attempting to use connectivity in the mental lexicon to predict various types of affective word covariates.

In Chapter 3, we investigated valence, arousal, and dominance, the three dimensions most commonly used when examining subjective meaning (see e.g., Osgood, 1975; Osgood, Suci, & Tannenbaum, 1957). A word's values on these factors are used in a wide variety of fields of research, to measure or control for the effect of emotional charge on some aspect. Traditionally, these ratings are gathered by asking participants to rate words one by one on each dimension, an expensive and exhaustive procedure. Recently, some researchers have proposed a computational approach to predict such ratings (e.g., Bestgen & Vincze, 2012; Mandera, Keuleers, & Brysbaert, 2015; Recchia & Louwerse, 2015). In short, their approach first derives similarity measures from word co-occurrences in text corpora, and subsequently estimates a word's affective values using its similarity towards words for which these values are already known. In Chapter 3 we adopted a similar technique, with the main difference that we obtained similarity indices from word association data rather than from text corpora. Using this method, we predicted the valence, arousal, and dominance of 14,000 words. We found that obtained estimates displayed very high correlations with human ratings; considerably higher, in fact, than corpus-based attempts at computationally estimating values on these dimensions. Taken together, we can conclude that these affective word covariates can successfully be predicted from word association data, and more generally, that affective meaning can be derived from a word's connectivity in the mental lexicon.

In Chapter 4, we examined whether it is possible to use word association data to predict the correspondence of words to various personality domains. This would be very useful as it would allow us to computationally derive personality from free-format descriptions of personality; currently, such descriptions have to be scored manually by experts, which takes a considerable amount of time and effort. To assess how well the technique described in Chapter 3 would function in this context, we first had participants complete a free-format self-description test, in which they described their own personality using any ten words. We then computationally encoded the correspondence of obtained responses to the Big Five personality factors, using their connectivity towards words for which this correspondence was already known. The same responses were also coded using expert ratings, which led to very similar personality profiles. As such, we conclude that one can predict a word's correspondence to personality domains from the word's semantic similarity to words for which this correspondence is already known. This is useful not only for encoding structured free-format personality measures such as the self-description test we investigated, but also allows for computationally estimating personality from unstructured sources of information such as posts on social media or diary entries.

Finally, in Chapter 5 we investigated brand personality, the set of human traits associated with a brand. Brand personality is generally measured through direct ratings, in which participant rate a brand's correspondence to various traits. In this chapter, we used a technique similar to that of Chapters 3 and 4 to assess the possibility of obtaining brand personality from the associations given to the brand in a word association task. If possible, this approach would hold two major advantages over direct ratings. First, in contrast to direct ratings, an association-based measure is very resistant against response bias. Second, this method only requires one session of data gathering (collecting associations towards the brand), after which a wide number of brand personality indices are available. Unfortunately, the obtained brand personality scores displayed a mixed correspondence to direct ratings; although correlations were very high for some dimensions, they were both low and nonsignificant for others. Of course, this does not necessarily indicate a problem with the association-based measure; indeed, one of the reasons we investigated this approach in the first place was that direct ratings are unsuited to measuring opinion as they are extremely sensitive to response bias. In order to obtain a better estimate of the performance of measuring brand personality through word associations, future research should include a much larger sample of brands, and compare obtained indices with more qualitative measures of brand personality.

6.2 Limitations and directions for future research

In this dissertation, we examined the relation between affective meaning and connectivity in the mental lexicon. We presented a number of findings that indicate that words tend to be connected to words with similar affective attributes, and that we can estimate a word's affective attributes from the words with which it is connected. At this point, we would like to address some of the limitations of the discussed research and offer a few suggestions for future investigation into this topic.

First, there is the issue of the representativity of participants. Although the word association dataset we made use of was constructed with the help of a very large number of participants (approximately 80,000) across all age-groups (ages between 7 and 96, M = 40), no information about their socio-economic status is available, making it hard to assess whether these participants can be considered a representative sample of the population. Additionally, part of the research in Chapter 4 and 5 was based entirely on first-year psychology students, clearly a skewed sample. As there is evidence that some aspects of the mental lexicon differ between people of different age (Entwisle, Forsyth, & Muuss, 1964; Palermo, 1971) and gender (Palermo & Jenkins, 1965), future research should investigate whether this is true for the role of affective meaning in the lexicon, as well as assess any effect of socio-economic background.

A related issue concerns language and culture: all studies were performed in Dutch, by Belgian or Dutch participants. Although for most of the effects we describe there is no a priori reason to expect a strong cultural influence, this is obviously not something we can rule out given the data at our disposal. As word association datasets are already freely available in many languages, including a number of non-Western languages, it should be relatively easy for future research to assess whether the findings we presented hold up across different languages and cultures.

Throughout this dissertation we made use of two extrapolation methods, each with a number of strengths and weaknesses. The first method, Orientation towards Paradigm Words, predicts a word's value on affective dimensions using that word's semantic similarity towards a number of paradigm words, words commonly used to describe extreme values on these dimensions. The second extrapolation method, *k*-Nearest Neighbors, predicts a word's score on some dimension using the mean score of its *k* nearest neighbors for which the value on that dimension is known. Comparing the two methods, the main advantage of the *k*-Nearest Neighbors method is a consistently higher accuracy (see Chapter 3 and 4), likely because predictions of this method are based directly on the human ratings with which accuracy is assessed. This does mean that this method can only predict values on dimensions for which ratings are already available for a number of words, which is definitely a large drawback. In contrast, the Orientation towards Paradigm Words method allows for estimates on any dimension, as long as a suitable choice of paradigm words can be found. We have seen the usefulness of this property in Chapter 5, where we made predictions on dimensions for which no direct ratings are available. As the only requirement of this method is choice of paradigm words, it seems likely this approach could also grant access to measures on more broad semantic domains or dimensions, such as male-female or young-old; this is definitely a topic that should be addressed in future research.

The two extrapolation methods were used to predict various affective word ratings with the help of word association data. For most affective dimensions, we find that both methods produce estimates that show a very high correspondence with human ratings. Clearly, the outlined technique is very successful at computationally generating affective word covariates. This approach does have one major drawback: it requires access to a large word association dataset. Although collecting word associations is a simple procedure, a considerable number of associations is needed in order to obtain a reasonable representation of the mental lexicon. Creating a dataset of this nature from scratch is definitely a very large undertaking; fortunately, as we have seen, these datasets are already freely available for a number of languages, including Dutch, English, French, German, Spanish, Rioplatense Spanish, Vietnamese, Japanese, and Cantonese. However, when working in languages where no sufficiently large word association datasets are available, one may want to consider looking into other sources of information about semantic relatedness, such as word co-occurrences in text corpora. An interesting possibility that has not been investigated as of yet is to combine text corpora with word association data in some way, such as using corpus-based indices to assist with creating or growing an association dataset, or perhaps integrating semantic information from both sources to obtain even richer data.

References

Bestgen, Y., & Vincze, N. (2012). Checking and bootstrapping lexical norms by means of word similarity indexes. *Behavior Research Methods*, *44*(4), 998–1006. doi:10.3758/s13428-012-0195-z

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 582–594. doi:10.1037//0278-7393.13.4.582

Cramer, P. (1968). *Word association*. New York, NY: Academic Press.

Entwisle, D., Forsyth, D., & Muuss, R. (1964). The syntactic-paradigmatic shift in children's word associations. *Journal of Verbal Learning and Verbal Behavior*, *3*(1), 19–29.

Klauer, K. C. (1997). Affective Priming. *European Review of Social Psychology*, *8*(1), 67–103. doi:10.1080/14792779643000083

Mandera, P., Keuleers, E., & Brysbaert, M. (2015). How useful are corpus-based methods for extrapolating psycholinguistic variables? *The Quarterly Journal of Experimental Psychology*, *68*(8), 1623–1642. doi:10.1080/17470218.2014.988735

Osgood, C. E. (1975). *Cross-cultural universals of affective meaning*. Urbana, IL: University of Illinois Press.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Palermo, D. S. (1971). Characteristics of word association responses obtained from children in grades one through four. *Developmental Psychology*, *5*(1), 118–123. doi:10.1037/h0031075

Palermo, D. S., & Jenkins, J. J. (1965). Sex Differences in Word Associations. *The Journal of General Psychology*, *72*, 77.

Pollio, H. (1964). Some semantic relations among word-associates. *The American Journal of Psychology*, *77*(2), 249–256. doi:10.2307/1420131

Recchia, G., & Louwerse, M. M. (2015). Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *Quarterly Journal of Experimental Psychology*, *68*(8), 1584–1598. doi:10.1080/17470218.2014.941296

Staats, A. W., & Staats, C. K. (1959). Meaning and m: correlated but separate. *Psychological Review*, *66*(2), 136–44.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. doi:10.1515/LANGCOG.2009.011